

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:25:14 ; Search time 14.4172 Seconds

(without alignments)
2831.539 Million cell updates/sec

Title: US-10-028-952a-10

Sequence: 1 IRHEGERTXEVXNKEAPG.....DXAIALQPKXRETLISQKK 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138.5	3.7	591	1	ALU1_HUMAN
2	126	3.4	591	1	ALU8_HUMAN
3	125	3.3	587	1	ALU2_HUMAN
4	123.5	3.3	593	1	ALU6_HUMAN
5	122	3.3	587	1	ALU3_HUMAN
6	110	2.9	2167	1	SHR1_RAT
7	108	2.9	593	1	ALU7_HUMAN
8	104	2.8	660	1	YH1L_EBV
9	100.5	2.7	585	1	ALU5_HUMAN
10	95	2.5	705	1	SYN1_HUMAN
11	93	2.5	450	1	A2AA_PIG
12	93	2.5	609	1	NMBL_HUMAN
13	92.5	2.5	1235	1	RMT1_ARATH
14	92	2.5	971	1	ARA1_GIBBR
15	89.5	2.4	1219	1	WIS1_YEAST
16	88	2.3	452	1	A2AA_BOVIN
17	88	2.3	706	1	SYN1_BOVIN
18	87	2.3	388	1	ROG_MOUSE
19	87	2.3	670	1	SYN1_MOUSE
20	86	2.3	391	1	ROG_HUMAN
21	86	2.3	450	1	A2AA_CANPO
22	86	2.3	450	1	A2AA_MOUSE
23	86	2.3	2424	1	CCAA_RABIT
24	85	2.3	833	1	SRC2_MOUSE
25	84	2.2	424	1	S3B4_HUMAN
26	84	2.2	514	1	VE2_HPV5B
27	84	2.2	870	1	SRC2_HUMAN
28	83	2.2	415	1	SYN1_CANPA
29	82.5	2.2	2483	1	PCX_DROME
30	82	2.2	704	1	SYN1_RAT
31	81.5	2.2	448	1	ACAC2_DICDI
32	81.5	2.2	696	1	JPH2_MOUSE
33	81.5	2.2	1076	1	VE38_SCHPO

ALIGNMENTS

RESULT 1
ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=95021756; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.",
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.",
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.",
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Miliuslavlevic A.;
RT "Reconstruction and analysis of human Alu genes.",
RL J. Mol. Evol. 32:105-121(1991).
RN [5]
RP MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
SENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
RN [6]
RP MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CAUTION: This Alu entry is provided in order to avoid the further
pollution of protein sequence databases with Alu-derived amino
acid sequences.
RN [7]
RP CAUTION: Alu repetitive sequences are interspersed in human and
primate genomes with an average spacing of 4 kb. Some of them are
actively transcribed by pol. III. Normal transcripts may contain
Alu-derived sequences in 5' or 3' untranslated regions. However,
cDNA libraries also contain partial and/or rearranged cDNAs
ligated with Alu-derived sequence in any orientation. Although Alu
elements (especially situated on the complementary strand) have a
great potential to create additional/alternative exons.

CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -1- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14567; -; NOT_ANNOTATED_CDS.
 CC -----
 CC Hypothetical protein.
 CC FT DOMAIN 1 96 FRAME-1.
 CC FT DOMAIN 100 195 FRAME-2.
 CC FT DOMAIN 199 294 FRAME-3.
 CC FT DOMAIN 298 393 FRAME-4.
 CC FT DOMAIN 397 492 FRAME-5.
 CC FT DOMAIN 496 591 FRAME-6.
 CC SQ SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;
 CC
 CC Query Match 3.7%; Score 138.5; DB 1; Length 591;
 CC Best Local Similarity 27.6%; Pred. No. 4.1e-05;
 CC Matches 48; Conservative 4; Mismatches 23; Indels 99; Gaps 2;
 CC
 CC QY 707 LESHVDPQPOHGGTSLQKNTKIX----- 731
 CC Db 121 LSPVRRDQPGHSTPDLQIKLQKLAGRGARLXSQLGLRQEDRLSPGRCSEPSRH 180
 CC QY 732 ----- 731
 CC Db 181 CTPWATERDPSVKKXXPGAVAHACNPSTLGGGRTYAGFEFTSLGNIVKPLKYK 240
 CC QY 732 ---PYLAHTGSLSYSEGGKXXIDXAQVEAAVAAXDAILQPGKERTLSQK 782
 CC Db 241 KMKGVVARRACSPSYSGKGRIRIKAEVEAAVSR-DRAIALQPGKERTLSQK 293
 CC
 CC RESULT 2
 CC ALU8_HUMAN STANDARD; PRT; 591 AA.
 CC ID ALU8_HUMAN
 CC AC P3195;
 CC DT 01-FEB-1995 (Rel. 31, Created)
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Alu subfamily 5X sequence contamination warning entry.
 CC OS Homo sapiens (Human).
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC CC NCBI_TaxId=9606;
 CC CC
 CC RN (1)
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=95021758; PubMed=7935834;
 CC RA Claverie J.-M.; Makalowski W.;
 CC RT "Alu alert";
 CC RL Nature 371:752-752(1994).
 CC RN [2]
 CC RP CONCEPT
 CC RX MEDLINE=92241891; PubMed=1572661;
 CC RA Claverie J.-M.;
 CC RT "Identifying coding exons by similarity search: alu-derived and other
 CC RT potentially misleading protein sequences.";
 CC RL Genomics 12:838-841(1992).
 CC RN [3]
 CC RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;
 RA Querlin Y.;
 RT "The Alu family developed through successive waves of fixation
 RT closely connected with primate lineage history.";
 RT J. Mol. Evol. 27:194-202(1988).
 CC
 CC [4]
 CC ALU FAMILIES CLASSIFICATION.
 CC RX MEDLINE=91178815; PubMed=1706781;
 CC RA Jurka J.; Milosavljevic A.;
 CC RT "Reconstruction and analysis of human Alu genes.";
 CC RL J. Mol. Evol. 32:105-121(1991).
 CC CC
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 CC REPEATS.
 CC
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC
 CC -1- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC
 CC -1- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC
 CC -1- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14574; -; NOT_ANNOTATED_CDS.
 CC -----
 CC Hypothetical protein.
 CC FT DOMAIN 1 96 FRAME-1.
 CC FT DOMAIN 100 195 FRAME-2.
 CC FT DOMAIN 199 294 FRAME-3.
 CC FT DOMAIN 298 393 FRAME-4.
 CC FT DOMAIN 397 492 FRAME-5.
 CC FT DOMAIN 496 591 FRAME-6.
 CC SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;
 CC
 CC Query Match 3.4%; Score 126; DB 1; Length 591;
 CC Best Local Similarity 26.3%; Pred. No. 0.00058;
 CC Matches 50; Conservative 5; Mismatches 33; Indels 102; Gaps 3;
 CC
 CC QY 691 LIPVIALXKXAGSLSRVDPQXGFTLSQKNTKIX----- 731
 CC Db 108 LTPVIPALMEED---HLRSVGRDQPGHGTPTSLKIQKLAGRGARLXSQLGLRQED 164
 CC QY 732 ----- 731
 CC Db 165 RINPGGGGSEPSRHCTPAWATERDSVKKXXRAEAGRGGRSLRSGHFGKRTXGQEF 224
 CC QY 732 ---PYLAHTGSLSYSEGGKXXIDXAQVEAAVAAXDAILQPG 772
 CC Db 225 ETSLANVPRILYKXKXKXPGVAVARACNPSSYSGKGRIRIATREAAVAVSR-DRAIALQ 283

```

0Y      773 GEXERTISOK 782
       7 | : ||| |
Db      284 GROSTPBOX 293

RESULT 3
ALU2_HUMAN
ID ALU2_HUMAN STANDARD; PRT; 587 AA.
AC P3189;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily SB sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euarcharia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RA MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1 MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1 MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1 CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences.
CC -1 CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -1 CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.

```

```

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14568; ; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN
FT 100 194 FRAME-1.
FT DOMAIN 198 292 FRAME-2.
FT DOMAIN 296 391 FRAME-3.
FT DOMAIN 395 489 FRAME-4.
FT DOMAIN 493 587 FRAME-5.
SQ SEQUENCE 587 AA; 63703 MW; 3EAB3B3E3929203 CRC64;
Query Match 3.3%; Score 125; DB 1; Length 587;
Best Local Similarity 43.4%; Pred. No. 0.00071;
Matches 43; Conservative 7; Mismatches 33; Indels 16; Gaps 3;
QY 691 LIPITALLXXKXGRSLESVRDQPHGHTL-----SLQKNTKTXPYLANTCSLSY 743
Db 6 LTPVITPLMEKEAGSGRGCEI-----ETLLANTYKPLRYKTKXPGVAGACSPSY 57
QY 744 SEGQGXIXDXQAEVEAAVREXDAXIALQPGXERETLSQK 782
Db 58 SGQXGEMAYTRFEALVSR-DRATLALQPGQSETPSQX 95
RESULT 4
ALU6 HUMAN STANDARD; PRT; 593 AA.
AC P39193;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily SP sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=8633509; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Miliusavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTRUCTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

```

CC REPEATS.

CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP

CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: This Alu entry is provided in order to avoid the further

CC pollution of protein sequence databases with Alu-derived amino

CC acid sequences.

CC -1- CAUTION: Alu repetitive sequences are interspersed in human and

CC primate genomes with an average spacing of 4 kb. Some of them are

CC actively transcribed by pol III. Normal transcripts may contain

CC Alu-derived sequences in 5' or 3' untranslated regions. However,

CC cDNA libraries also contain partial and/or rearranged cDNAs

CC ligated with Alu-derived sequence in any orientation. Although Alu

CC elements (especially situated on the complementary strand) have a

CC great potential to create additional/alternative exons,

CC consideration should be given to the possibility that the presence

CC of an Alu in an open reading frame may have resulted from a

CC cloning artifact or may be due to misinterpretation of sequencing

CC data. This point has been overlooked on several occasions, with

CC the consequence of erroneous Alu-derived amino acid sequences

CC being reported.

CC -1- CAUTION: Any significant similarity of a putative protein sequence

CC with an Alu-translated entry must be taken as a warning that a

CC part of Alu repeat may have been artifactually included in the

CC coding nucleotide sequence.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC DR EMBL: U14572; -; NOT ANNOTATED CDS.

CC KM Hypothetical protein.

CC FT DOMAIN 1 97 FRAME-1.

CC FT DOMAIN 101 196 FRAME-2.

CC FT DOMAIN 200 295 FRAME-3.

CC FT DOMAIN 299 395 FRAME-4.

CC FT DOMAIN 494 593 FRAME-5.

CC FT DOMAIN 593 64603 MW; 136EP344ACD12A2 CRC64;

CC SQ SEQUENCE

CC Query Match 3.3%; Score 123.5; DB 1; Length 593;

CC Best Local Similarity 25.7%; Pred. No. 0.00099;

CC Matches 45; Conservative 7; Mismatches 23; Indels 100; Gaps 3;

CC -----

CC QY 707 LESHVDQPYQHGHTLSLQKNTKIX----- 731

CC Db 122 LRSGVRQDPQHGTTPSLKIQKLAGRGACIKSQLGRLENRLNPGGGCGEPRSRH 181

CC QY 732----- 731

CC Db 182 CTPAMATRAKLRLKXXXXPAAVAHACNPSTLGGRGGRITXGREFTSLTMEKPLYYKX 241

CC QY 732 ---PYVAHTCSLSYSEKXGXIDXAQEVAAAVRXDXAIALOPG-XERTTSSQK 782

CC Db 242 KMXGCVAAHACNPFYSGXGRIRIAXTEAAVAISR-DRAIALPQGGGRNSVSKK 295

CC RESULT 5

CC ALU3 HUMAN STANDARD; PRT; 587 AA.

CC ID ALU3 HUMAN STANDARD; PRT; 587 AA.

CC AC P39190;

CC DT 01-FEB-1995 (Rel. 31, Created)

CC DT 01-FEB-1995 (Rel. 31, Last sequence update)

CC DT 10-OCT-2003 (Rel. 42, Last annotation update)

CC DE Alu subfamily SBI sequence contamination warning entry.

CC OS Homo sapiens (Human).

CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

CC CC NCBI_Taxid=9606;

CC RN [1]

RP SEQUENCE FROM N.A.

RP MEDLINE=95021758; PubMed=7935834;

RP RA Claverie J.-M., Makalowski W.;

RP RT "Alu alert."

RP RL Nature 371:752-752(1994).

RP RN [2]

RP RP CONCEPT.

RP RX MEDLINE=92241891; PubMed=1572661;

RP RA Claverie J.-M.;

RP RT "Identifying coding exons by similarity search: alu-derived and other

RP RT potentially misleading protein sequences."

RP RL Genomics 12:838-841(1992).

RP RN [3]

RP RP ALU FAMILIES CLASSIFICATION.

RP RX MEDLINE=88333009; PubMed=3138422;

RP RA Quentin Y.;

RP RT "The Alu family developed through successive waves of fixation

RP RT closely connected with primate lineage history."

RP RL J. Mol. Evol. 27:194-202(1988).

RP RN [4]

RP RP ALU FAMILIES CLASSIFICATION.

RP RX MEDLINE=91178815; PubMed=1706781;

RP RA Jurka J., Milosavljevic A.;

RP RT "Reconstruction and analysis of human Alu genes."

RP RL J. Mol. Evol. 32:105-121(1991).

CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE

CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING

CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

CC REPEATS.

CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP

CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: This Alu entry is provided in order to avoid the further

CC pollution of protein sequence databases with Alu-derived amino

CC acid sequences.

CC -1- CAUTION: Alu repetitive sequences are interspersed in human and

CC primate genomes with an average spacing of 4 kb. Some of them are

CC actively transcribed by pol III. Normal transcripts may contain

CC Alu-derived sequences in 5' or 3' untranslated regions. However,

CC cDNA libraries also contain partial and/or rearranged cDNAs

CC ligated with Alu-derived sequence in any orientation. Although Alu

CC elements (especially situated on the complementary strand) have a

CC great potential to create additional/alternative exons,

CC consideration should be given to the possibility that the presence

CC of an Alu in an open reading frame may have resulted from a

CC cloning artifact or may be due to misinterpretation of sequencing

CC data. This point has been overlooked on several occasions, with

CC the consequence of erroneous Alu-derived amino acid sequences

CC being reported.

CC -1- CAUTION: Any significant similarity of a putative protein sequence

CC with an Alu-translated entry must be taken as a warning that a

CC part of Alu repeat may have been artifactually included in the

CC coding nucleotide sequence.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC DR EMBL: U14569; -; NOT ANNOTATED CDS.

CC KM Hypothetical protein.

CC FT DOMAIN 1 96 FRAME-1.

CC FT DOMAIN 100 194 FRAME-2.

CC FT DOMAIN 198 292 FRAME-3.

CC FT DOMAIN 296 391 FRAME-4.

CC FT DOMAIN 395 489 FRAME-5.

CC FT DOMAIN 493 587 FRAME-6.

CC SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DFF235 CRC64;

CC Query Match 3.3%; Score 122; DB 1; Length 587;

Best Local Similarity 44.6%; Pred. No. 0.0013;
Matches 41; Conservative 8; Mismatches 41; Indels 2; Gaps 2;
QY 691 LIPVIALXXAGSLESRVDQXQHGSETSLQKNTKIXVLTATCSLSYSGXGX 750
DB 6 LIPVLPALWEAEKSGSGOIEITIPAKTVKP-RLYKRYKXKPGVAGACSPSLGKXGR 64
QY 751 IDKAEVEAAVAXDXAIALPOGXESETLSOK 782
DB 65 MAXTRAEALAVSR-DEPTALPOGROSETPSQK 95
RESULT 6
SHKL RAT
ID SHKL RAT STANDARD; PRT; 2167 AA.
AC 09WV48; 09WV48; 09WV48; 09WV48;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
DLG4.
RC TISSUE=Brain;
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hixio K., Deguchi M., Ide N., Takeuchi M., Takai Y.,
RT "Synamon, a novel neuronal protein interacting with synapse-associated
protein 90/postsynaptic density-95-associated protein.",
RL J. Biol. Chem. 274:27463-27466(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=99360650; PubMed=10433268;
RA Naibit S., Kim E., Tu J.C., Xiao B., Sala C., Valschanoff J.,
RT Weisberg R.J., Morley P.F., Sheng M.,
RT "Shank, a novel family of postsynaptic density proteins that binds to
the NMDA receptor/PSD-95/GKAP complex and cortactin.",
RL Neuron 23:569-582(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobben S., Suedhof T.C., Stahl B.,
RT "The G protein-coupled receptor C11 interacts directly with proteins
of the Shank family.",
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naibit S., Yoon U., Hwang J.I., Suh P.G., Sheng M., Kim E.,
RT "Characterization of the Shank family of synaptic proteins. Multiple
genes, alternative splicing, and differential expression in brain and
development.",
RL J. Biol. Chem. 274:29510-29518(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Heenck H.-H., Baechner D., Richter D., Kriegenkamp H.-U.,
RT "Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.",
RL J. Biol. Chem. 274:33997-33001(1999).
RN [6]
RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.

RX MEDLINE=99360651; PubMed=10433269;
RA Tu J.C., Xiao B., Naibit S., Yuan J.P., Petralia R.S., Brakeman P.,
RA Doan A., Akalu V.K., Danahan A.A., Sheng M., Morley P.F.,
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
postsynaptic density proteins.",
RL Neuron 23:583-592(1999).
RN [7]
RP INTERACTION WITH SPTAN1.
RX MEDLINE=21523912; PubMed=11509555;
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
RA Buck F., Richter D., Gundelfinger E.D., Kriegenkamp H.-U.,
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
multidomain Shank protein family interact with the cytoskeletal
protein alpha-fodrin.",
RL J. Biol. Chem. 276:40104-40112(2001).
RN [8]
RP FUNCTION.
RX MEDLINE=21389514; PubMed=11498055;
RA Sala C., Plech V., Wilson N.R., Passafium M., Liu G., Sheng M.,
RT "Regulation of dendritic spine morphology and synaptic function by
Shank and Homer.",
RL Neuron 31:115-130(2001).
RN [9]
RP REVIEW.
RX MEDLINE=20267867; PubMed=10860936;
RA Sheng M., Kim E.,
RT "The Shank family of scaffold proteins.",
RL J. Cell Sci. 113:1851-1856(2000).
CC -1- FUNCTION. Seems to be an adaptor protein in the postsynaptic
density (PSD) of excitatory synapses that interconnects receptors
of the postsynaptic membrane including NMDA-type and metabotropic
glutamate receptors, and the actin-based cytoskeleton. May play a
role in the structural and functional organization of the
dendritic spine and synaptic junction. Overexpression promotes
maturation of dendritic spines and the enlargement of spine heads
via its ability to recruit Homer to postsynaptic sites, and
enhances presynaptic function.
CC -1- SUBUNIT: May homodimerize via its SAM domain. Interacts with
SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
the PDZ domain (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=5;
Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9WV48-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9WV48-2; Sequence=VSP_006072; VSP_006073;
CC Name=3;
CC IsoId=Q9WV48-3; Sequence=VSP_006074;
CC Name=4; Synonym=A;
CC IsoId=Q9WV48-4; Sequence=VSP_006075;
CC Name=5;
CC IsoId=Q9WV48-5; Sequence=VSP_006076; VSP_006077;
CC -1- TISSUE SPECIFICITY: Expressed only in brain (neurofil of cortex,
CAL region hippocampus and molecular layer of cerebellum).
CC -1- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
to high levels at 3-4 weeks before dropping slightly in adulthood.
CC Expressed in the cortex and the molecular layer of the cerebellum
at postnatal day 7. Isoform 2 expression does not change during
development of both cortex and cerebellum. Isoform 4 expression
decreases significantly during development of cortex but not
cerebellum.
CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC -1- SIMILARITY: Contains 7 ANK repeats.
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -1- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been artifactually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14573; -, NOT_ANNOTATED_CDS.
 CC KW Hypothetical protein.
 CC FT DOMAIN 1 97 FRAME-1.
 CC FT 101 196 FRAME-2.
 CC FT 200 295 FRAME-3.
 CC FT 239 395 FRAME-4.
 CC FT 399 484 FRAME-5.
 CC FT 488 593 FRAME-6.
 CC SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;
 CC -----
 CC Query Match 2.9%; Score 108; DB 1; Length 593;
 CC Best Local Similarity 24.6%; Pred. No. 0.027;
 CC Matches 43; Conservative 7; Mismatches 25; Indels 100; Gaps 3;
 CC -----
 CC QY 707 LESRRVDPQXGHEISLQKNTKIX----- 731
 CC DB 122 ILSGVADQPGHGETPLTKIQKLAGRGGLXSQLGLRQENLNPQGGGCEPDRSH 181
 CC QY 732 ----- 731
 CC DB 182 CTPANATRAKLLKXXXXGAVAHANCFETLGGRGWITXGQEFETSLANNVKKPLYYKY 241
 CC QY 732 ---PYVLAHTGSLSTSEKGGXXIDXAQVEBAAVKDXAIALQPG-XERETLSQK 782
 CC DB 242 KXVPGVAGACNPSYSGXGRIRAXTREAEVAVSR-DEATALQPGQGBRNSVSKK 295
 CC -----
 CC RESULT 8
 CC YH11_EBV STANDARD; PRT; 660 AA.
 CC ID YH11_EBV
 CC AC P03181;
 CC DT 21-JUL-1986 (Rel. 01, Created)
 CC DT 21-JUL-1986 (Rel. 01, Last sequence update)
 CC DE 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE Hypothetical BHLF1 protein.
 CC OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 CC OC Gammaherpesvirinae; no RNA stage; Herpesviridae;
 CC OC Gammapherpesvirinae; Lymphocryptoviruses.
 CC NCBI_TaxID=10377;
 CC RX [1]
 CC RP SEQUENCE FROM N.A.
 CC RA MEDLINE=84270667; PubMed=6087149;
 CC RA Baer R.T., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 CC Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 CC Tufnell P.S., Barrett B.G.,
 CC RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome";
 CC RL Nature 310:207-211 (1984).
 CC -----

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; V01555; -, NOT_ANNOTATED_CDS.
 CC DR PIR; A03742; OQHE3.
 CC KW Hypothetical protein; Early protein; Repeat.
 CC FT DOMAIN 149 648 4 X 125 AA TANDM REPEATS.
 CC FT REPEAT 149 273 1.
 CC FT REPEAT 274 398 2.
 CC FT REPEAT 399 523 3.
 CC FT REPEAT 524 648 4.
 CC SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;
 CC -----
 CC Query Match 2.8%; Score 104; DB 1; Length 660;
 CC Best Local Similarity 23.2%; Pred. No. 0.07;
 CC Matches 87; Conservative 13; Mismatches 103; Indels 172; Gaps 21;
 CC -----
 CC QY 18 APGAP-PVSPRGARQGXRRPCG-----PVKYHYSDRXXTDPVRRGGEPR- 61
 CC DB 188 APGAPGPAP-PGCGAAVPSGATPHERSGADPPA-----AARLPFERQ--EPRL 238
 CC QY 62 -GALAXAKRPAPRRPATPSGASAPWGYLIGVYTOQTXQVSGELREDN-----LPWX 116
 CC DB 239 PQLDAAQRCPCAG--PEPTSGAA-----QTHRRPPGCPRSANRNGCCRTYR 285
 CC QY 117 RRAKARLLIFSTWTDXESGASRFPFGFXAXXVKVTGTLGLMRPSXHSDAFXSD 176
 CC DB 286 RR-----SGAQGHPPPG-----AQGRSGPTG-GRPAPG----- 315
 CC QY 177 VGSYHAKAEFTKRWIVHPLIGXSNVYTVRQVSTLXMCCHGNPQCYENRXXRH 236
 CC DB 316 -----APGTPA----- 321
 CC QY 237 VVYLXGANGAKXXSVGLXLNASKSESPPGRTIRQRGASVGLGXFPRLSP----- 288
 CC DB 322 ---APPGGAIVPG---ATPHPERGSPADPPAARLPERQEPRLPQLDAAQRC 373
 CC QY 289 PAGRPPPTXXRAGGVPRRPAQ-----PSSXACPSSEWF-----TERGRK 330
 CC DB 374 PAG--PPPTSGAAAGQTHRRPPGCPRSANPG---CPRTWRRSQAQGRHPPGAGOR 427
 CC QY 331 -----GGXPLARRAP 340
 CC DB 428 PSCPTGRRPAPAPAP 442
 CC -----
 CC RESULT 9
 CC ALUS_HUMAN STANDARD; PRT; 585 AA.
 CC ID ALUS_HUMAN
 CC AC P39192;
 CC DT 01-FEB-1995 (Rel. 31, Created)
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Alu subfamily SC sequence containing warning entry.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 CC RX [1]
 CC RP SEQUENCE FROM N.A.
 CC RA MEDLINE=95021758; PubMed=7935834;
 CC RA Claverie J.-M., Makalowski W.,
 CC RT "Alu alert";
 CC RL Nature 371:752-752 (1994).
 CC RX [2]
 CC RP CONCEPT.
 CC RA MEDLINE=92241891; PubMed=1572661;
 CC -----

RA Clavette J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RT Genomics 12:838-841(1992).
 RN [3]
 RN ALU FAMILIES CLASSIFICATION.
 RP MEDLINE=88333009; PubMed=3138422;
 RX Quercin Y.;
 RA "The Alu family developed through successive waves of fixation
 RT closely connected with primate lineage history.";
 RT J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RN ALU FAMILIES CLASSIFICATION.
 RP MEDLINE=91178815; PubMed=1706781;
 RX Jurka J., Mironavljovic A.;
 RA "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 CC REPEATS.
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -1- CAUTION: This Alu entry is provided in order to avoid the further
 CC pollution of protein sequence databases with Alu-derived amino
 CC acid sequences.
 CC -1- CAUTION: Alu repetitive sequences are interspersed in human and
 CC primate genomes with an average spacing of 4 kb. Some of them are
 CC actively transcribed by pol III. Normal transcripts may contain
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,
 CC cDNA libraries also contain partial and/or rearranged cDNAs
 CC ligated with Alu-derived sequence in any orientation. Although Alu
 CC elements (especially situated on the complementary strand) have a
 CC great potential to create additional/alternative exons,
 CC consideration should be given to the possibility that the presence
 CC of an Alu in an open reading frame may have resulted from a
 CC cloning artifact or may be due to misinterpretation of sequencing
 CC data. This point has been overlooked on several occasions, with
 CC the consequence of erroneous Alu-derived amino acid sequences
 CC being reported.
 CC -1- CAUTION: Any significant similarity of a putative protein sequence
 CC with an Alu-translated entry must be taken as a warning that a
 CC part of Alu repeat may have been attractually included in the
 CC coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U14571; ; NOT ANNOTATED_CDS.
 KW Hypothetical protein.
 FT DOMAIN 1 95 FRAME-1.
 FT DOMAIN 2 99 193 FRAME-2.
 FT DOMAIN 3 197 291 FRAME-3.
 FT DOMAIN 4 295 389 FRAME-4.
 FT DOMAIN 5 393 487 FRAME-5.
 FT DOMAIN 6 491 585 FRAME-6.
 FT SEQUENCE 585 AA; 63957 MW; 46E88C4F493650A7 CRC64;
 SQ
 Query Match 2.7%; Score 100.5; DB 1; Length 585;
 Best Local Similarity 41.4%; Pred. NO. 0.13; Indels 17; Gaps 4;
 Matches 41; Conservative 6; Mismatches 35;
 Oy 691 LIPVILALXXAKAGRSLSRRDQPKXHGELT-----SLQKNTKIXPYVLAHTCSLSY 743
 Db 6 LTPVILPLWKEKAGSGSGEI-----ETILANVWKPRLYKKYKXKLGVAVRACSPSY 57
 Oy 744 SEGKXGXIDXAQGEVLAANVXKDXALALQPKKERETLSQK 782

Db 58 SGCGXGRIMXTRAEVAVSR-DEATLQPG-DRARLRXK 94
 RESULT 10
 SYN1 HUMAN
 ID SYN1 HUMAN STANDARD; PRT; 705 AA.
 AC P17600; O75825;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Synapsin I (Brain protein 4.1).
 GN SYN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=90243651; PubMed=2110562;
 RA Suedhof T.C.;
 RT "The structure of the human synapsin I gene and protein.";
 RL J. Biol. Chem. 265:7849-7852(1990).
 RN [2]
 RP SEQUENCE OF 1-125 FROM N.A.
 RX MEDLINE=90368667; PubMed=2118519;
 RA Sauerwald A., Hoesehe C., Oeschwald R., Kiliann M.W.;
 RT "The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and
 RT CAAT-less, phylogenetically conserved sequence with cell
 RT type-specific promoter function.";
 RL J. Biol. Chem. 265:14932-14937(1990).
 RN [3]
 RP SEQUENCE OF 1-258 FROM N.A.
 RA Griefham D.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,
 CC binds to the cytoskeleton, and is believed to function in the
 CC regulation of neurotransmitter release.
 CC -1- SUBCELLULAR LOCATION: Synapse.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Synapsin IA;
 CC IsoId=P17600-1; Sequence=Displayed;
 CC Name=Synapsin IB;
 CC IsoId=P17600-2; Sequence=VSP 006316, VSP 006317;
 CC -1- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
 CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
 CC SYNAPSIN I IN THE NERVE TERMINAL.
 CC -1- SIMILARITY: Belongs to the synapsin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M58378; AAC41930.1; JOINED.
 CC EMBL: M58321; AAC41930.1; JOINED.
 CC EMBL: M58341; AAC41930.1; JOINED.
 CC EMBL: M58351; AAC41930.1; JOINED.
 CC EMBL: M58353; AAC41930.1; JOINED.
 CC EMBL: M58359; AAC41930.1; JOINED.
 CC EMBL: M58371; AAC41930.1; JOINED.
 CC EMBL: M58372; AAC41930.1; JOINED.
 CC EMBL: M58373; AAC41930.1; JOINED.
 CC EMBL: M58374; AAC41930.1; JOINED.
 CC EMBL: M58375; AAC41930.1; JOINED.
 CC EMBL: M58376; AAC41930.1; JOINED.
 CC EMBL: M58377; AAC41930.1; JOINED.
 CC EMBL: M58378; AAC41931.1; ALT_SEQ.

DR EMBL, M58321, AAC41931.1, JOINED.
 DR EMBL, M58341, AAC41931.1, JOINED.
 DR EMBL, M58351, AAC41931.1, JOINED.
 DR EMBL, M58359, AAC41931.1, JOINED.
 DR EMBL, M58371, AAC41931.1, JOINED.
 DR EMBL, M58372, AAC41931.1, JOINED.
 DR EMBL, M58373, AAC41931.1, JOINED.
 DR EMBL, M58374, AAC41931.1, JOINED.
 DR EMBL, M58375, AAC41931.1, JOINED.
 DR EMBL, M58376, AAC41931.1, JOINED.
 DR EMBL, M58377, AAC41931.1, JOINED.
 DR EMBL, M58301, AAC60508.1, JOINED.
 DR EMBL, AL009172, CAAL5657.1, JOINED.
 DR PIR, A35363, A35363.
 DR HSSP, P17599, IADUX.
 DR Genew, HGNC:11494, SYN1.
 DR MIM, 313440, -.
 DR GO, GO:0005803, C:secretory vesicle; TAS.
 DR GO, GO:0005215, F:transporter activity; TAS.
 DR GO, GO:0006899, P:nonselective vesicle transport; TAS.
 DR GO, GO:0007268, P:Synaptic transmission; TAS.
 DR InterPro, IPR001359, Synapsin.
 DR Pfam, PF02078, Synapsin, 1.
 DR Pfam, PF02750, Synapsin, C, 1.
 DR PRINTS, PR01368, SYNAPSIN.
 DR PROSITE, PS00415, SYNAPSIN_1, 1.
 DR PROSITE, PS00416, SYNAPSIN_2, 1.
 DR Synapse, Phosphorylation; Neurone, Repeat; Actin-binding;
 Alternative splicing.
 FT DOMAIN 1 28 A.
 FT DOMAIN 29 112 B (LINKER).
 FT DOMAIN 113 420 C
 (ACTIN-BINDING AND SYNAPTIC-VESSICLE BINDING).
 FT DOMAIN 421 655 D (PRO-RICH LINKER).
 FT MOD_RES 9 9 E.
 FT MOD_RES 568 568 PHOSPHORYLATION (BY PKA AND CAMK1).
 FT MOD_RES 605 605 PHOSPHORYLATION (BY CAMK2).
 FT VAASPLIC 661 669 NKSQSLRYA--> KASPAQAQ (in isoform Synapsin IB).
 FT VAASPLIC 670 705 /FTid=VSP_006316.
 FT COMPACT 138 138 Missing (in isoform Synapsin IB).
 FT SEQUENCE 705 AA, 73954 MW, 487831123FF6882F CAC64;
 Query Match 2.5%, Score 95, DB 1, Length 705;
 Best Local Similarity 27.5%, Pred. No. 0.51;
 Matches 53, Conservative 16, Mismatches 64, Indels 60, Gaps 11;
 OY 177 VGSSVHXAETKRWIVHPLIGXKS-----WVTVVQVAFLLIMCCCHGNPAQYERN 230
 DB 388 VGSSN-----PLIGDHODEDKOLIVLVNKKMAAL-----PRGQRD 425
 OY 231 RXRRLVVLGXGANGAKXSVGLXNKSSESXPXGTTIRORRAASVGLGXEP-----R 285
 DB 426 ASP-----GRSGHGQTPSPGALPLGRQTSQ-QPAGPPAQRPPPGG-GPPQPGPPOR 476
 OY 286 LSPF-AGRPP-----STRXXRAGGRVRAFGGXAACPPSSMET-----GRGRKG-- 331
 DB 477 QGPFQGGPPPGGCHSLGPPRAGSPFLPQLPSTLAPQGPADQAAAPPPGGGRGRSPV 536
 OY 332 ---GXPLARHP 340
 DB 537 AGGPPAPPARPP 549
 RESULT 11
 AAA_PIG STANDARD; PRT; 450 AA.
 ID A2AA_PIG
 AC P18671;

DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Alpha-2A adrenergic receptor (Alpha-2A receptor) (Alpha-2AR).
 GN ADR2A OR A2AR
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 OX [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 212-232.
 RC TISSUE=Liver;
 RX MEDLINE=91009167, PubMed=2170371;
 RA Guyer C.A., Horstman D.A., Wilson A.L., Clark J.D., Kravoe E.J. Jr.,
 RA Limbird L.E.;
 RT "Cloning, sequencing, and expression of the gene encoding the porcine
 RT alpha 2-adrenergic receptor. Allosteric modulation by Na+, H+, and
 RT amiloride analogs.";
 RT J. Biol. Chem. 265:17307-17317(1990).
 RN [2]
 RP MUTAGENESIS OF CYS-442.
 RX MEDLINE=93216775, PubMed=8385131;
 RA "Mutations of the alpha 2A-adrenergic receptor that eliminate
 RT detectable palmitoylation do not perturb receptor-G-protein
 RT coupling.";
 RT J. Biol. Chem. 268:8003-8011(1993).
 CC - FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-
 CC induced inhibition of adenylyl cyclase through the action of G
 CC proteins.
 CC - SUBCELLULAR LOCATION: Integral membrane protein.
 CC - MISCELLANEOUS: Alpha2-adrenergic receptor shows an allosteric
 CC modulation by Na(+), H(+) and amiloride analogs.
 CC - SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL, J05652, AAA30984.1, -.
 CC PIR, A38316, A38316.
 CC HSSP, P29274, 1MMH.
 CC InterPro, IPR000276, GPCR_Rhodopsn.
 CC Pfam, PF00001, 7tm.1, 1.
 DR PRINTS, PR00237, GPCR_RHODOPSN.
 DR PROSITE, PS00237, G-PROTEIN_RECEP_F1_1, 1.
 DR PROSITE, PS00262, G-PROTEIN_RECEP_F1_2, 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;
 KM Multigene family; Phosphorylation; Lipoprotein; Palmitate.
 FT DOMAIN 1 33
 FT TRANSMEM 34 59 1 (POTENTIAL).
 FT DOMAIN 60 70 2 (POTENTIAL).
 FT TRANSMEM 71 96 3 (POTENTIAL).
 FT DOMAIN 97 106 4 (POTENTIAL).
 FT TRANSMEM 107 129 5 (POTENTIAL).
 FT DOMAIN 130 149 6 (POTENTIAL).
 FT TRANSMEM 150 173 7 (POTENTIAL).
 FT DOMAIN 174 192 8 (POTENTIAL).
 FT TRANSMEM 193 217 9 (POTENTIAL).
 FT DOMAIN 218 374 10 (POTENTIAL).
 FT TRANSMEM 375 399 11 (POTENTIAL).
 FT DOMAIN 400 430 12 (POTENTIAL).
 FT TRANSMEM 431 450 13 (POTENTIAL).
 FT CARBOHYD 10 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 14 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 106 188 188 BY SIMILARITY.
 FT LIPID 442 442 5-palmitoyl cysteine.
 FT SITE 113 113 IMPLICATED IN LIGAND BINDING (BY

FT	SITE	200	200
FT	SITE	204	204
FT	MUTAGEN	442	442
FT	SEQUENCE	450 AA;	48975 NM; 7907DSB4737074E CRC64;
SO	Query Match	Best Local Similarity 24.3%; Pred. No. 0.48;	Matches 43; Conservative 14; Mismatches 74; Indels 46; Gaps 8
OY	DGANGAKKXSVGLNINASKSSRXGITRQRGRASVGLGXPRXRSPAGRPPEPTRXX	300	
DB	GVRGVGAEEAPLPVLQNGAFPAPAQP----	RDLND-GLDLEBSSESHERRPGRPBE	312
OY	RA-----GGRVPPR---ADPGSKACPSWETGRGRKGXP.LRRHAPHV	343	
DB	RGPRAKSARASGVKEGDPLPRGPCGAFGCAPAA-----TGAGEERG-	-VA	357
OY	AALAEFKXSSTIHRHTSACIFMXIIIILFM-----VDIQKWDC--XTMXYFMP	391	
DB	KSRMRGRQRNRKRFFFLVALAVIGVFVCWPFPFFTITLIRVGCSVPFITLFKEFFWF	414	

RESULT 12
NMBL_HUMAN ID_NMBL_HUMAN STANDARD; PRT; 609 AA.

```

OYQ9Y6RO;
DT16-OCT-2001(Ref., 40, Created)
DT16-OCT-2001(Ref., 40, Last sequence update)
DT15-MAR-2004(Ref., 43, Last annotation update)
DE Numb.-like protein (Numb-R).
GN NMBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mamalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
EN SEQUENCE FROM N.A.
RP MEDLINE=97447799; PubMed=9303539;
RA Salcini A.E., Confalonieri S., Doria M., Santolini E., Tassi E.,
BA Minenkova O., Cesarati G., Pellicci P.G., Di Fiore P.P.;
RT "Binding specificity and in vivo targets of the EH domain, a novel
RL protein-protein interaction module."
RU Gene Dev. 11:2239-2249(1997).
[2]
EN SEQUENCE OF 9-609 FROM N.A.
RP TISSUE=Eye;
RC MEDLINE=22386257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheefel C.F., Bat N.K.,
RA Hopkins R.F., Urdan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stopleton M.W., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Tohiyuki S., Abramoen R.D., Mullaly S.J.,
RA Raab S.S., Loggiani N.A., Peters G.J., Abramson R.D., Mulihay S.J.,
RA Bosak S.A., McGean P.T., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahuy U., Helton E., Ketteman W., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard C.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimmwood J., Schmutz V., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Stallus D.E.,
RA Scherzer A., Schön J.E., Jones S.J.M., Martz M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
CC -!- SUBUNIT: Associates with EphA2 and NOTCH1.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Contains 1 PID domain.
```

Query Match	Best Local Similarity	Score 93:	DB 1:	Length 609:
Matches	80: Conservative	25: Mismatch	140: Indels	112: Gaps
CC	EMBL; AF015041; AAD01549.1; ALT_INIT.			
CC	EMBL; BC001794; AA001794.1; ALT_INIT.			
CC	HSSP; P16554; 2NMJ.			
CC	DR Genew; HGNC:8061; NUMBL.			
CC	DR MIM; 604018; -			
CC	DR GC; CC0007399; P:neurogenesis; TAS.			
CC	DR InterPro; IPR006020; PTB_PTD.			
CC	DR Pfam; PF00640; PTD; 1.			
CC	DR SMART; SMO0462; PTB; 1.			
CC	DR PROSITE; PS01179; PTD; 1.			
CC	DOMAIN 74 223 PTD.			
CC	FT 427 446 POLY-GLN.			
CC	FT 609 AA; 64891 MW; 69458D161B60FB4 CRC64;			
CC	SEQUENCE			
CC	Query Match	2.5%: Score 93:	DB 1:	Length 609:
CC	Best Local Similarity	22.4%: Pred. No. 0.67:		
CC	Matches	80: Conservative	25: Mismatch	140: Indels
CC	CC	EMBL; AF015041; AAD01549.1; ALT_INIT.		
CC	CC	EMBL; BC001794; AA001794.1; ALT_INIT.		
CC	CC	HSSP; P16554; 2NMJ.		
CC	CC	DR Genew; HGNC:8061; NUMBL.		
CC	CC	DR MIM; 604018; -		
CC	CC	DR GC; CC0007399; P:neurogenesis; TAS.		
CC	CC	DR InterPro; IPR006020; PTB_PTD.		
CC	CC	DR Pfam; PF00640; PTD; 1.		
CC	CC	DR SMART; SMO0462; PTB; 1.		
CC	CC	DR PROSITE; PS01179; PTD; 1.		
CC	CC	DOMAIN 74 223 PTD.		
CC	CC	FT 427 446 POLY-GLN.		
CC	CC	FT 609 AA; 64891 MW; 69458D161B60FB4 CRC64;		
CC	CC	SEQUENCE		
CC	CC	Query Match	2.5%: Score 93:	DB 1:
CC	CC	Best Local Similarity	22.4%: Pred. No. 0.67:	Length 609:
CC	CC	Matches	80: Conservative	25: Mismatch
CC	CC	EMBL; AF015041; AAD01549.1; ALT_INIT.		
CC	CC	EMBL; BC001794; AA001794.1; ALT_INIT.		
CC	CC	HSSP; P16554; 2NMJ.		
CC	CC	DR Genew; HGNC:8061; NUMBL.		
CC	CC	DR MIM; 604018; -		
CC	CC	DR GC; CC0007399; P:neurogenesis; TAS.		
CC	CC	DR InterPro; IPR006020; PTB_PTD.		
CC	CC	DR Pfam; PF00640; PTD; 1.		
CC	CC	DR SMART; SMO0462; PTB; 1.		
CC	CC	DR PROSITE; PS01179; PTD; 1.		
CC	CC	DOMAIN 74 223 PTD.		
CC	CC	FT 427 446 POLY-GLN.		
CC	CC	FT 609 AA; 64891 MW; 69458D161B60FB4 CRC64;		
CC	CC	SEQUENCE		
CC	CC	Query Match	2.5%: Score 93:	DB 1:
CC	CC	Best Local Similarity	22.4%: Pred. No. 0.67:	Length 609:
CC	CC	Matches	80: Conservative	25: Mismatch
CC	CC	EMBL; AF015041; AAD01549.1; ALT_INIT.		
CC	CC	EMBL; BC001794; AA001794.1; ALT_INIT.		
CC	CC	HSSP; P16554; 2NMJ.		
CC	CC	DR Genew; HGNC:8061; NUMBL.		
CC	CC	DR MIM; 604018; -		
CC	CC	DR GC; CC0007399; P:neurogenesis; TAS.		
CC	CC	DR InterPro; IPR006020; PTB_PTD.		
CC	CC	DR Pfam; PF00640; PTD; 1.		
CC	CC	DR SMART; SMO0462; PTB; 1.		
CC	CC	DR PROSITE; PS01179; PTD; 1.		
CC	CC	DOMAIN 74 223 PTD.		
CC	CC	FT 427 446 POLY-GLN.		
CC	CC	FT 609 AA; 64891 MW; 69458D161B60FB4 CRC64;		
CC	CC	SEQUENCE		
CC	CC	Query Match	2.5%: Score 93:	DB 1:
CC	CC	Best Local Similarity	22.4%: Pred. No. 0.67:	Length 609:
CC	CC	Matches	80: Conservative	25: Mismatch
CC	CC	EMBL; AF015041; AAD01549.1; ALT_INIT.		
CC	CC	EMBL; BC001794; AA001794.1; ALT_INIT.		
CC	CC	HSSP; P16554; 2NMJ.		
CC	CC	DR Genew; HGNC:8061; NUMBL.		
CC	CC	DR MIM; 604018; -		
CC	CC	DR GC; CC0007399; P:neurogenesis; TAS.		
CC	CC	DR InterPro; IPR006020; PTB_PTD.		
CC	CC	DR Pfam; PF00640; PTD; 1.		
CC	CC	DR SMART; SMO0462; PTB; 1.		
CC	CC	DR PROSITE; PS01179; PTD; 1.		
CC	CC	DOMAIN 74 223 PTD.		
CC	CC	FT 427 446 POLY-GLN.		
CC	CC	FT 609 AA; 64891 MW; 69458D161B60FB4 CRC64;		
CC	CC	SEQUENCE		
CC	CC	Query Match	2.5%: Score 93:	DB 1:
CC	CC	Best Local Similarity	22.4%: Pred. No. 0.67:	Length 60

RT Sequence features of the regions of 1,367,185 bp covered by 19
 RT physically assigned pi and TAC clones."
 CC DNA Res. 5:203-216(1998).
 CC -1- FUNCTION: Eliminates the production of nonsense-containing RNAs
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: Belongs to the DNA/NAM7 helicase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB013394; BAB10240.1; -
 KM Hypohectical protein: Hydrolase; Helicase; ATP-binding; Zinc-finger.
 FT ZN_FING 148 178 C2H2-TYPE (ATYPICAL) (POTENTIAL).
 FT ZN_FING 202 232 C4-TYPE (Potential).
 FT NP_BIND 507 514 ATP (POTENTIAL).
 FT DOMAIN 102 105 POLY-SER.
 SQ SEQUENCE 1235 AA; 134870 MW; 17733F6C524D2E6A CRC64;
 Query Match 2.5%; Score 92.5; DB 1; Length 1235;
 Best Local Similarity 29.9%; Pred. No. 1.6;
 Matches 44; Conservative 9; Mismatches 59; Indels 35; Gaps 9;
 QY 226 QYERNR---XHLVYVLGXGANGAKXSVGLXNLSKSESR-----PXGTRIQRG 274
 Db 929 QGCKPRKIXNDRLRFYGGAGMIGNDNPFSG--NPNDNRGSGRAGSYLPBG---PNEG 984
 QY 275 ASVGL---GXPRR--LSPAGRPPTSTXXRAGRVPRRAGPSXACEPSWETGR 329
 Db 985 ASPPGMPAGYPIPRVPLSPFGPPSPQPYA-----IPTRGP-VGAVPHAPQPGNHGCGA 1037
 QY 330 KGGXPLARAPVRAAEFXXSSTIHN 356
 Db 1038 GGGTSVGGHLPHQQA-----TQHN 1056
 RESULT 14
 AREA GIBFU STANDARD; PRT; 971 AA.
 ID AREA GIBFU STANDARD; PRT; 971 AA.
 AC P78688;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nitrogen regulatory protein area.
 GN AREA.
 OS *Gibberella fujikuroi* (Bakanease and foot rot disease fungus) (*Fusarium*
 OS moniliforme).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Nectriaceae; *Gibberella*.
 OX NCBI_TaxID=5127;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ms67;
 RX MEDLINE=99168774; PubMed=10071216;
 RA Tudyanski B., Homann Y., Feng B., Marzluf G.A.;
 RA "Isolation, characterization and disruption of the area nitrogen
 RA regulatory gene of *Gibberella fujikuroi*.";
 RL Mol. Genet. 261:106-114(1999).
 CC -1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
 CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: Contains 1 GATA-type zinc finger.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Y11006; CA71897.1; -
 DR HSSP; P17429; 4GAT.
 DR InterPro; IPR000679; Znf GATA.
 DR Pfam; PF00320; GATA; 1.
 DR PRINTS; PR00619; GATAZNFINGER.
 DR SMART; SM00401; Znf GATA; 1.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
 KM Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KM Nuclear protein; Nitrate assimilation.
 FT ZN_FING 694 718 GATA-TYPE.
 SQ SEQUENCE 971 AA; 103580 MW; 887D882141C7453 CRC64;
 Query Match 2.5%; Score 92; DB 1; Length 971;
 Best Local Similarity 20.8%; Pred. No. 1.4;
 Matches 76; Conservative 30; Mismatches 138; Indels 122; Gaps 18;
 QY 6 GEETEXYXNKK--EAPGAPVSPRGAGGXRRCGPVKYHSDRXXTDPVRGGEPRG- 62
 Db 580 GTTDFPDNNGDMESNGLERSQGSFRGMLRQHPXL-----PRNASTVHFGGQGNF 634
 QY 63 -ALAXGKRAARRPPATRSGXARMGVXIGRTCTGXQVSGELREDNLPWXRARA 121
 Db 635 EQLAQSQSSPADDGNQTMGFS--VAPSR-----PSPFMSQGS- 674
 QY 122 RLILIFSTNDXSGASRSFPFGFAXXVYRKTTGTGLMRPSKSDVAFSFDGSSV 181
 Db 675 -----TTNIGAAAGNND-----GNAFTTCTNCTFTPLMRNNPBGQPLCNAAGFLK 724
 QY 182 FXEAEFTKMIYHPLIGKXSWVTVVRQVSTLLKMCCHGAPAQYERNRKXHLVYVLG 241
 Db 725 HG-----VVRPL-----SLKTDVTK-----KRNK----- 744
 QY 242 XGANGAKXSVGLXNLSKSESRPXGTRIQRGASVGLX-----PXPL-SP 288
 Db 745 SGTN---VPGV--GSSTSKTASTLNRKNTLSMSTATNSTKPNSSNTPTVTP 797
 QY 289 PAGRPSPS-----TRXXRAGRVPR--APGGSXACEPSSWETGRRGXGP 334
 Db 798 PATSQPSSKDVDSPVSGTTSAGTASHTNSHFSGPSSGAV-----GKGWVP 848
 QY 335 LASHAP 340
 Db 849 IAAAP 854
 RESULT 15
 MYSS YEAST STANDARD; PRT; 1219 AA.
 ID MYSS YEAST STANDARD; PRT; 1219 AA.
 AC O04439;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE *Myosin-5* isoform.
 GN MYO5 OR YMR109W OR YMR718.08.
 OS *Saccharomyces cerevisiae* (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.A.,
 RA Unger K., Dye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Batteil B.G.;
 RA "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome
 RT XIII."
 RL Nature 387:90-93(1997).

```

RN SUBUNIT, AND MUTAGENESIS OF TRP-11123.
RX MCHLINA-2189311; PubMed-1190111;
RT "The novel adaptor protein, Wtldp, and Vtldp, a homolog of
RT Wiskott-Aldrich syndrome protein-interacting protein (WIP), may
RT antagonistically regulate type I myosin in Saccharomyces
RT cerevisiae."
RL Genetics 160:923-934(2002).
CC -1- SUBUNIT: Binds via its SH3 domain to BBCL.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
-----
DR EMBL; Z49702; CAA89745.1; -.
DR PIR; S54570; S54570.
DR HSSP; P08799; IMND.
DR GeneOntology; 142777; -.
DR SCD; S0004715; MYOS.
DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomyces); IDA.
DR GO; GO:0007047; P:cell wall organization and biogenesis; IMP.
DR GO; GO:0006887; P:exocytosis; IMP.
DR GO; GO:0007121; P:polar budding; IMP.
DR GO; GO:0009551; P:saltinility response; IGI.
DR InterPro; IPR0000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00181; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD0000355; myosin_head; 1.
DR ProDom; PD0000066; SH3; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR MYOSIN; Coiled coil; ATP-binding; SH3 domain.
KT MYOSIN; Coiled coil; ? MYOSIN HEAD-LIKE.
FT DOMAIN 1
FT DOMAIN 1085 1147 SH3.
FT DOMAIN 1190 1218 COILED COIL (POTENTIAL).
FT NP BIND 129 136 ATP (POTENTIAL).
FT DOMAIN 1011 1016 POLY-PRO.
FT DOMAIN 1060 1063 POLY-ALA.
FT DOMAIN 1073 1081 POLY-PRO.
FT DOMAIN 1204 1218 ASP/GLU-RICH (ACIDIC).
FT MUTAGEN 1123 1123 W->S; ABOLISHES INTERACTION WITH BBCL.
SQ SEQUENCE 1219 AA; 13689 MW; DFB9SECI6B1CD29 CRC64;

Query Match 2.4%; Score 89.5; DB 1; Length 1219;
Best Local Similarity 26.8%; Freq. No. 2.9; Indels 27; Gaps 4;
Matches 30; Conservative 11; Mismatches 44;

256 LNASKESRPFKXITQR-----GASVGLGEXPRISPPAGRPPESTRXXRRAGGEVP 307
1035 LTASGSMRNPSPPTATATPATPATPAAMAGSGROA-NIPPPPPPPSSKKEPFEAA 1093
308 RAAPRGGAAC-----PSMWR-----GRKGXPLAHAPH 341
1094 YDPSSGSSSELPLKKGDVITYITREPPSGMSGLKLDGSKESWVFATMKH 1145

```

Search completed: April 27, 2004, 10:53:39
Job time : 16.4172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:45:34 ; Search time 41.7603 Seconds

(without alignments)
5923.479 Million cell updates/sec

Title: US-10-028-952a-10

Perfect score: 3752
Sequence: 1 IRHEGERTXEVNCKEAPG.....DXAIALPGKXERETLSQKX 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL.25.*
2: sp_archaea.*
3: sp_bacteria.*
4: sp_fungi.*
5: sp_human.*
6: sp_invertebrate.*
7: sp_mammal.*
8: sp_mhc.*
9: sp_organelle.*
10: sp_phage.*
11: sp_plant.*
12: sp_rodent.*
13: sp_virus.*
14: sp_vertebrate.*
15: sp_unclassified.*
16: sp_virus.*
17: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	3.2	543	16 Q82AU7	Q82AU7 streptomyc
2	117	3.1	95	16 Q84BM1	Q84BM1 bacillus me
3	113	3.0	77	16 Q8OCV6	Q8OCV6 straphylococ
4	104.5	2.8	539	10 Q8XHV3	Q8XHV3 oryza sativ
5	103	2.7	376	16 Q8XYP2	Q8XYP2 deinosoccus
6	102	2.7	62	17 Q972FO	Q972FO sulfolobus
7	99.5	2.7	655	16 Q9FBR7	Q9FBR7 streptomyc
8	99.5	2.6	674	2 Q9FBR7	Q9FBR7 streptomyc
9	97.5	2.6	377	2 Q91426	Q91426 mycobacteri
10	97.5	2.6	750	10 Q9SDP5	Q9SDP5 oryza sativ
11	97.5	2.6	2496	11 Q8VHD8	Q8VHD8 mus musculu
12	96.5	2.6	542	12 Q84357	Q84357 mastomys na
13	96	2.6	552	16 Q8RSH9	Q8RSH9 deinosoccus
14	95.5	2.5	160	6 Q8HKB4	Q8HKB4 macaca faec
15	95	2.5	936	10 Q9XKX5	Q9XKX5 oryza sativ
16	94.5	2.5	409	10 Q8S156	Q8S156 oryza sativ

17	94	2.5	461	10 Q9FTD4	Q9FTD4 oryza sativ
18	93	2.5	362	10 Q8S031	Q8S031 oryza sativ
19	93	2.5	494	17 Q9HNR5	Q9HNR5 halobacteri
20	93	2.5	1305	11 Q80YF9	Q80YF9 mus musculu
21	92.5	2.5	439	10 Q8R226	Q8R226 oryza sativ
22	92.5	2.5	1243	10 Q8S3K7	Q8S3K7 arabidopsis
23	92	2.5	264	10 Q7XGB8	Q7XGB8 oryza sativ
24	91.5	2.4	437	5 Q8SXL9	Q8SXL9 drosophila
25	91.5	2.4	1627	10 Q84ZL0	Q84ZL0 oryza sativ
26	91	2.4	216	2 Q8VPR2	Q8VPR2 micrococcus
27	91	2.4	390	6 Q97GE9	Q97GE9 smitopsis
28	90.5	2.4	272	10 Q7XGE9	Q7XGE9 oryza sativ
29	90.5	2.4	298	2 Q8XVMS	Q8XVMS thermus the
30	90.5	2.4	437	5 Q9VWNS	Q9VWNS drosophila
31	90.5	2.4	788	16 Q9SL52	Q9SL52 rhizobium 1
32	90	2.4	529	2 P94909	P94909 microbacter
33	90	2.4	601	5 Q62185	Q62185 caenorhabdi
34	90	2.4	3262	11 Q9EQJ5	Q9EQJ5 mus musculu
35	89.5	2.4	751	10 Q9FRM7	Q9FRM7 arabidopsis
36	89.5	2.4	759	2 Q68843	Q68843 streptomyc
37	89.5	2.4	776	10 Q8H078	Q8H078 arabidopsis
38	89.5	2.4	780	10 Q93ZT6	Q93ZT6 arabidopsis
39	89.5	2.4	780	10 Q8H179	Q8H179 arabidopsis
40	89	2.4	690	10 Q7XDU3	Q7XDU3 oryza sativ
41	88.5	2.4	219	12 Q91T11	Q91T11 tupiaia herp
42	88.5	2.4	224	17 Q9YD60	Q9YD60 aeropyrum P
43	88.5	2.4	457	10 Q94LX1	Q94LX1 oryza sativ
44	88.5	2.4	846	16 Q82FV2	Q82FV2 streptomyc
45	88	2.3	172	4 Q8NF08	Q8NF08 homo sapien

ALIGNMENTS

RESULT 1
Q82AU7 PRELIMINARY; PRT; 543 AA.
ID Q82AU7;
AC Q82AU7;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV5958.
OS Streptomycetes avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCPI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinohe M., Takahashi Y., Horikawa H., Nakazawa H., Osomoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomycetes
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RC MEDLINE=22608306; PubMed=12692562;
RX Ikeda H., Ishikawa J., Hanamoto A., Shinohe M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomycetes avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005044; BAC73670.1; .
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 543 AA; 60464 MW; E59CE37E88AD0B05 CRC64;
Query Match 3.2%; Score 119; DB 16; Length 543;
Best Local Similarity 22.1%; Pred. No. 0.00092;

Matches 96; Conservative 41; Mismatches 145; Indels 152; Gaps 24;

QY 26 PRGARGXRRRCGPVKYHS-----DXXYD-----PARGGEPG 62
 Db 121 PLREKGGRRAGG---YHSQEWHAARLHLEDRLEQARTDRDAGVHVVRGK--- 173
 QY 63 ALAAGAKRPAARRPGATSGXSARW-----GVXLRGYTCQTXQVSKGE----- 106
 Db 174 HLAIRHHLAAQAQ---LLETGWRQWRBAERWFLKADGSGKRYGNETVRVSPBGEVASIKLP 231
 QY 107 --LREDNLN-----KKRAKAILIFSTNTXESGASGSPFPG 146
 Db 232 APLKELNAPRGRVYLAQVHFPRGEMADREYANALVRYHYTARG--RWYLTASN 289
 QY 147 XAXXVRKY-----TTGITGL-----WRPSXK---SDVAFKSPVY-GSSYXEA 186
 Db 290 QIPRTQVPIPAALQGVIGVDMADHLAMRLDVHGNPVGDRPRRPFPLSGPAGHRDAQ 349
 QY 187 FTKRWIVHPLIGXSW-----VXTVVRQVSPFLIMCCHGNPAQYERNRERXHLVYVL 240
 Db 350 -----VHALTGLHMAACGVTAAVEDLDFADKTRERHG-----RKRFRROLISGW 398
 QY 241 GXGANGAKXXSV-----GLXINA-----SKSESRP-XGTIR--QSRGASVGL-- 279
 Db 399 PTGRLRRLTSMADHTGIAIAYDPATYTSRKAQHMQKPLTSKIKRTTRDAAVAIGR 458
 QY 280 --GXPRPLSPAGRPPTSTXXRAGRVPR-----RAFPGSKAECPSSWE 324
 Db 459 AOGPIRRRT---APPHDQSDRAGHRTVQARPGIPWRBGTBPRILGPRTRVGP--- 510
 QY 325 TGRGRKGGXPLARH 338
 Db 511 -GRGANAGDQNAQH 523

RESULT 2

Q848W1 PRELIMINARY; PRT; 95 AA.

AC Q848W1. 24. Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Bacillus megaterium.
 OG Plasmid pBM400.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=1404;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=QM B1551;
 RA MEDLINE=21150449; PubMed=11251820;
 RA Kunimatsuya M., Stevenson D.M., Zhou Y., Vary P.S.;
 RT "Analysis of the replication region and identification of an RNA operon
 on pBM400 of Bacillus megaterium QM B1551.";
 RL Mol. Microbiol. 39:1010-1021(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=QM B1551;
 RA Vary P.S., Scholte M.D., White C.A., Kunimatsuya M.;
 RT "Complete Sequencing and Characterization of pBM400 from Bacillus
 megaterium.";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF142677; AA052805.1; --
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR009019; KI_pick.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 95 AA; 10201 MW; 3FD616C2E7379256 CRC64;

Query Match 3.1%; Score 117; DB 2; Length 95;
 Best Local Similarity 41.3%; Pred. No. 0.0018;
 Matches 31; Conservative 1; Mismatches 42; Indels 0; Gaps 0;
 136 GARSSEFPFGXAXXVRKVTGTTGTLWRPSXHSDDVAFXSFDVGSSTXHEAEFTKRWIVHP 195

Db 14 GLSDPVYHGRALIAQRKATPGITGLSPPRVHIDGFWHLVDVSGSHPGAVVGPKNMAVRP 73
 QY 196 LIGXSWXTVVRQ 209
 Db 74 LKRYASWQVNVVRQ 87

RESULT 3

Q8CQV6 PRELIMINARY; PRT; 77 AA.

AC Q8CQV6. 23. Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Conserved hypothetical protein.
 GN SE2364.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcaceae.
 CX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qian Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016751; AA005906.1; --
 DR InterPro; IPR009019; KI_pick.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 77 AA; 8402 MW; 9ED757600933581A CRC64;

Query Match 3.0%; Score 113; DB 16; Length 77;
 Best Local Similarity 43.3%; Pred. No. 0.0038;

Matches 29; Conservative 1; Mismatches 37; Indels 0; Gaps 0;

QY 143 PRGYAXXVRKVTGTTGTLWRPSXHSDDVAFXSFDVGSSTXHEAEFTKRWIVHPLIGXSW 202
 Db 3 PHGRAIAQRKATPGITGLSPPRVHIDGFWHLVDVSGSHPGAVVGPKNMAVRP LKRYASW 62
 QY 203 VXTVVRQ 209
 Db 63 VQNVVRQ 69

RESULT 4

Q9XHV3 PRELIMINARY; PRT; 539 AA.

AC Q9XHV3. 12. Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DE 10A191.13.
 GN 10A191.13.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Gramineae; Oryzoideae; Oryza.
 CX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Lemont;
 RA Vysotskaya V.S., Schwartz J.R., Osborne B.I., Wing R., Yu G., Kwan A.,
 RA Liu S., Lee J., Tortum M., Luros J., Li J., Kremenetskaya I., Oji O.,
 RA Theologis A.;
 RT "Oryza sativa chromosome 1 BAC 10A191.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007858; AAD39598.1; --
 DR Gramene; Q9XHV3; --
 SQ SEQUENCE 539 AA; 58285 MW; D30F0686A5946605 CRC64;

Query Match 2.8%; Score 104.5; DB 10; Length 539;
 Best Local Similarity 22.0%; Pred. No. 0.003;
 Matches 95; Conservative 39; Mismatches 181; Indels 117; Gaps 17;

QY 4 EHGEEITXEVXNCKEAPGAPVPGARAGGXRPPGCG-PPVKYHYSDRXXTDP-----VRR 56
 DB 120 EEDENTE-----EPPRAPIVSPRAGPRGARLADPPACPH--ARPTTPRIVLVYLR 172
 QY 57 GGEPRGALLAKGAKRPAARRPGATRGKSRMGVXIGRYTCOTXXOVXG-----105
 DB 173 PPRPAPAAAY-TRALKKERNHARGSGKADMLAEKRCRVPCNCRSAGFPFHGLSVE 231
 QY 106 -----ELREDPNLPWXRRAKARLILFSTNTDSEGASRFXPPGFXXXVAKV 154
 DB 222 IFNHLPIYDVCALNTDRHIFSSRTAFIEIRLI-----SNTYPAHWLGRRLKET 280
 QY 155 TTGITGLWRPSXHSDFVAFKSPDVGSYH-----XEAFTKRWIVH-----PLIGXXSW 202
 DB 281 DTMLEARSARMQEMEA-----TASESHLPWWLCPSPAHDCRCCTMHAHSGPWWGSSYT 335
 QY 203 VXTV-VROVSTFLLMCCCH-----GNPAQYENNRXRLVYVLGANGAKOXKXVGLXLN 257
 DB 336 VITLQPLSLATSSGLGTHTRARGERGBGSRPAG-----GGGDSGGGEEVGATV 389
 QY 258 A-----SKSESPPXGTRRGRGASVGLGXFP 284
 DB 390 AVPAGQSAATNLRSLSLASHGLGSEERRRPREETRGDSRP--PRQRHSGARLHRRSP 447
 QY 285 RISPAGPPPTSTRXKXAGRVPRAPPGGSAECPPSSWET-GRGRKGXHLAHPHYR 343
 DB 448 RVGAAPLGPFAARQI-----WPPRSFGPAAAVTPPGPATMAAARRQIWPFRPSPVG 502
 QY 344 ARAEFXXSSTIH 355
 DB 503 AAAPLGPVAVH 514

RESULT 5

Q9RYP2 PRELIMINARY; PRT; 376 AA.

AC Q9RYP2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Adenine deaminase-related protein.
 GN DRA0268.
 OS Deinococcus radiodurans.
 CC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 CC Deinococcaceae; Deinococcus.
 CX NCBI_TaxID=1259;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567265;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crebly M., Shen M.,
 RA Vamathavan J., Lam P., McDonald L., Uetrecht T., Zaleski C.,
 RA Makarova K.S., Aavand L., Daly M.J., Minchin K.W., Fleischman R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1."
 RL Science 286:1571-1577 (1999).
 DR EMBL; AB001863; AAF12376.1; -.
 DR TIGR; C75580; C75580.
 DR Complete proteome.
 SQ SEQUENCE 376 AA; 39845 MW; 7AB7FF32F8C45651 CRC64;

Query Match 2.7%; Score 103; DB 16; Length 376;
 Best Local Similarity 23.0%; Pred. No. 0.031; Indels 136; Gaps 18;
 Matches 82; Conservative 15; Mismatches 123; Indels 136; Gaps 18;
 QY 20 GAPVSPGARGGXRP--CGPPVYHYSDRXXTDPVRGGEPRGAL---AXGAKRPAAR 74

DB 64 GRPAPGAGQSGAGCEPRVYAHPLY---SGGRADRAPGASAGARYREILGPAPRPGD 120
 QY 75 RQATRS-GXSARMGVXIGRYTCOTXXOVXGELR-----EDRNLP--KXRRAKRL 123
 DB 121 RFWVSRLRAAALATLRG-----AGULRGSGSAARRAPPAKRRRRRAKL 169
 QY 124 ILFSTNTDSEGASRFXPPGFXAXKVRKVTGIGLWRPSXHSDFVAFKSPDVGSYHX 183
 DB 170 -----GRGLHS-----ARALHADVSRPDH-----192
 QY 184 EAEFTKRWIVHPLIGXXSWXIVVROVSTFLLMCCCHG---NEAQYENNRXRLVYVL 240
 DB 193 -----PACAGGRRRPAGRRRPLRARRVVELL 219
 QY 241 GXGANGKXXSVGLXINASKSESPXGTRRGRGAS-----VGLGXFPRL 286
 DB 220 DVGGRPARRHP-GHQSARASGSGPAR--RQRGRPARGSGARTARRRHPGRGRGPPA 276
 QY 287 SPAPGPPPTSTRXKXAGRVPRAPPGGSAECPPSSWETGRGRKGXPLA--RHAP 340
 DB 277 VAPALR--GSDDRPASGR-GRRCPPGPD-----GGGALAGLHLSRHP 318

RESULT 6

Q972F0 PRELIMINARY; PRT; 62 AA.

AC Q972F0;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Hypothetical protein STS132.
 GN STS132.
 OS Sulfolobus tokodaii.
 CC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 CC Sulfolobus.
 CX NCBI_TaxID=111955;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / ?;
 RX MEDLINE=21456156; PubMed=1572479;
 RA Kwaratabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takemitsu M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Kuhsaki J., Kuhsaki N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Yamagishi A.,
 RA Ohima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermoacidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7."
 RL DNA Res. 8:123-140 (2001).
 DR EMBL; AP000985; BAB6219.1; -.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 62 AA; 6269 MW; 134E8A389A075DD8 CRC64;

Query Match 2.7%; Score 102; DB 17; Length 62;
 Best Local Similarity 48.1%; Pred. No. 0.0044; Indels 0; Gaps 0;
 Matches 25; Conservative 2; Mismatches 25; Indels 0; Gaps 0;

QY 152 RKYTTGITGLWRPSXHSDFVAFKSPDVGSYHXAFTKRWIVHPLIGXXSW 203
 DB 11 RKYTPGITSSRARAHIDPVCYIDVSSHPGGAAPKGRARPLMGVSVMV 62

RESULT 7

Q9FBR7 PRELIMINARY; PRT; 655 AA.

AC Q9FBR7;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Putative iron chelataase.
 GN SC05278 OR SC0512.02.
 OS Streptomyces coelicolor.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN (1)
 RP SOURCE FROM N.A.
 RC STRAIN=J3(2) / M15;
 RX MEDLINE=2196410; PubMed=1200953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittschke E., Rajandream M.A., Rutherford K., Ruter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierzbicki A., Woodward J., Barrill B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147 (2002).
 DR EMBL AL939123; CAC04493.1;
 KW Complete proteome.
 SQ SEQUENCE 655 AA; 67324 MW; 523242CD93B7F109 CRC64;

Query Match 2.7%; Score 99.5; DB 16; Length 655;
 Best Local Similarity 24.1%; Pred. No. 0.14;
 Matches 99; Conservative 33; Mismatches 144; Indels 135; Gaps 25;

QY 4 EHG---EETXEVNKKXKAPAP--VSPRGAR--GGRRPCGPVYKHYSDXKTDV 54
 DB 103 EHGITYRTTTPGPLETDDADGAPLVVVDLALSLVAGNR---AAVQLGAD---VYTV 155
 QY 55 RRGEPFGALAYGAKPAPARRPQATRSXGSAFVGLGRYTC-QTXXQVXGELREDNTL 113
 DB 156 ERGG--PRRTV---RPRARWLAVCRSADVGR---VSRHLIDFEVAFLSAAGR---L 201
 QY 114 PMKBRKARLLIFSTNTXESGASRSPFPGFAXYKATGTTGTLMRPSXSDVAF 173
 DB 202 PGRELLA---ETGPRPEAGAA---AGEGELPAGLPAARAKAAAGV--R 244
 QY 174 SPDVSSYHKEAEFTKMTVHPILIGXSMVXTVAVQVFTLL--XMCCHGNPAQYERNR 231
 DB 245 PVPVGDAAIGRA---RELIGPDAG---VRELLALRLARTLAALHGRPS--- 287
 QY 232 RKRLVYVVGXGANGKXKXVGLX---NASKSESPX---GTIORRGAS 276
 DB 288 -----ASADDSDAAALILGIVAPQAPGRPSGPPRPPAPGPTGLPTVERRAAG 339
 QY 277 VGLKAPXR-----LSPAGRPPTSTKXKRAAGRPV---RR 309
 DB 340 -GDGRFRPDGDRGLDTGPAGIGTAPGAAPSPADGSPYEDDEPAAARDTVPLSGPRR 398
 QY 310 APGPGXAECPSSWETGRGKXGKXPLA---RRAPIVRAARAFXXSSTIHR 357
 DB 399 STGPVST-----RGTVIGTRARDLRIALIVRIVR---AAVHQR 435

RESULT 8
 Q9FLJ6 PRELIMINARY; FRT; 674 AA.
 AC Q9FLJ6;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Putative HRP.
 GN PUTATIVE HRP.
 OS Acidovorax avenae subsp. avenae.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Comamonadaceae; Acidovorax.
 OX NCBI_TaxID=80870;
 RN (1)
 RP SOURCE FROM N.A.
 RC STRAIN=PAVE18101;
 RA Takikawa Y., Bo S., Adachi S., Kojima M.;
 RT "Deduced hrp genes from Burkholderia glumae and Acidovorax avenae

RT subsp. avenae.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL AB053454; BAB20912.1;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:000306; F:protein secretion; IEA.
 DR InterPro; IPR001712; Bact_FHIPP.
 DR InterPro; IPR006302; HrcV.
 DR Pfam; PF00771; FHIPP; 1.
 DR PRINTS; PR00949; TYPE3IMPAPROT.
 DR TIGRFAMs; TIGR01399; hrcv; 1.
 DR PROSITE; PS00994; FHIPP; 1.
 SQ SEQUENCE 674 AA; 73214 MW; 73AB86ED21CBA3C CRC64;

Query Match 2.7%; Score 99.5; DB 2; Length 674;
 Best Local Similarity 28.5%; Pred. No. 0.15;
 Matches 45; Conservative 9; Mismatches 61; Indels 43; Gaps 7;

QY 226 QYERNRRKRLHLYVLGKXG-----ANGAKXXSVGLXINAKSESRP----- 265
 DB 275 QUTRTREMYLAGLTLGFAIPGFPVAPPCLLAAGL---VGGGQWLSSRSARPPRAQ 330
 QY 266 -----XGTRORRGASVGLKAPXR-----LSP--AGRPSTRTXKRAAGRPVPR 309
 DB 331 DFGQARLPAPARRQGR--GHGHPAPRARIQAAVDPVVSARPPAQCRARPGPGQTR 368
 QY 310 APGPGXAECPSSWETGRGKXGKXPLARHAPHYARAE 347
 DB 369 RAADAAGALPRRRDVGGRPPG--PALRHHRHARRAK 425

RESULT 9
 Q9LA26 PRELIMINARY; FRT; 377 AA.
 AC Q9LA26;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Gsa protein.
 GN GSA.
 OS Mycobacterium avium.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1764;
 RN (1)
 RP SOURCE FROM N.A.
 RX MEDLINE=20569498; PubMed=1118730;
 RA Bull T.C., Sheridan J.M., Martin H., Sumar N., Tizard M.,
 RA Hermon-Taylor J.;
 RT "Further studies on the Gsa element: A novel mycobacterial insertion
 RT sequence (IS1612), inserted into an acetylase gene (mpa) in
 RT Mycobacterium avium subsp. silvaticum but not in Mycobacterium avium
 RT subsp. paratuberculosis";
 RL Vet. Microbiol. 77:453-463 (2000).
 DR EMBL AJ251811; CAB93583.1;
 SQ SEQUENCE 377 AA; 41141 MW; BCC5CD72A7843880 CRC64;

Query Match 2.6%; Score 97.5; DB 2; Length 377;
 Best Local Similarity 23.8%; Pred. No. 0.12;
 Matches 69; Conservative 16; Mismatches 94; Indels 111; Gaps 14;

QY 18 APCAPVSPR--GAGGXRPPGPPVKYHYSD-----RXXTDVRSGEPRGALAYGAKR 70
 DB 72 APGDEPDRRAGDAGRRTRPRRVPHRSATPATIRANGVYRGLDVGIC--GAAR 129
 QY 71 PAARPP--GATRSXGAPRWGVXGRYT-----CQTXXQVXGELREDNTL---PKXR 117
 DB 130 RRARRPAGHGAGSGFSRTGRVSATFPAPSRIRFARTNORSPGVTVDDVLDLWFW-- 187
 QY 118 RAKARLLIFSTNTDXXS-----GASRSF-----XPF 144
 DB 188 -----FMHSIDRSDDPAVAVSVEALSNGPPTDLMGYKNNLVGLDGKPLEPRRY 238
 QY 145 GFAXXVYKRVTTGIT-----GLMRPSXSD 169

```

Db      239 GYMPFMRKFFLLGATYAAHQATFFGASIVAKLGGYDIDFGLDQLFYRAALIRBVTID 298
QY      170 VAFKSPDV---GSSYHAEAF---TKRWIVH---PLIG-XXSWVTVVRQ 209
Db      299 RVCVCDPVGPGSGTQPIREHYRLRLMDLHGDYPLGGRVSWAYLRVVE 348

RESULT 10
ID      Q9SDP5      PRELIMINARY;      PRT;      750 AA.
AC      Q9SDP5;
DT      01-MAY-2000 (TREMBlrel. 13, Created)
DT      01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE      Hypothetical protein.
OS      Oryza sativa (Rice).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzeae; Oryza.
OX      NCBI_TaxID=4530;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Nipponbare;
RA      Sasaki T., Matsumoto T., Yamamoto K.;
RT      "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT      clone:P0038P12."
RL      Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF008036; BAA88189.1; -.
DR      Gramene; Q9SDP5; -.
KW      Hypothetical protein.
SQ      SEQUENCE 750 AA; 82191 MW; C96A063FF1FE2E83 CRC64;

Query Match      2.6%; Score 97.5; DB 10; Length 750;
Best Local Similarity 27.3%; Pred. No. 0.27; Mismatches 60; Indels 17; Gaps 5;
Matches 35; Conservative 16;

Db      236 LVYVLGXGANGAKXXSVGLXNLSKSESRPXGTI---RQRGASVGLGXPRLSPPAGR 292
QY      59 LVVVLGGGSGSGMEERAG--SMEAQEPNSAAKKTLMPPRSRGVADRSIFGRSHPPRQ 117
Db      293 PPSTXXXRAGCGVPPRAPGPGSGXACPSWETGRGRKXGPIARAPHVRAAEYXXS 352
QY      118 PPPPPASRQLAAR-HRQAPP-----PAS-----HRQAPPERRRSGAALASGHSSA 164
Db      353 TTHRRATS 360
QY      165 TVHRRNST 172

RESULT 11
ID      Q8VHD8      PRELIMINARY;      PRT;      2496 AA.
AC      Q8VHD8;
DT      01-MAR-2002 (TREMBlrel. 20, Created)
DT      01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Hornerin.
GN      HORNERIN OR 110033K191R1K.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ICR;
RA      MDL:LINE=21601613; PubMed=11572870;
RA      Makino T., Takashi M., Morohashi M., Huh N.-H.;
RT      "Hornerin, a Novel Profilaggrin-like Protein and Differentiation-
RT      specific Marker Isolated from Mouse Skin."
RL      J. Biol. Chem. 276:47445-47452(2001).
CC      -1- SIMILARITY: BELONGS TO THE S-100 FAMILY.
DR      EMBL; AY027660; AAK15791.1; -.

```

```

DR      MGD; MGI:1915973; Hornerin.
DR      GO; GO:0001533; C:cornified envelope; IDA.
DR      InterPro; IPR001751; CAP_S100.
DR      InterPro; IPR002048; EF-hand.
DR      Pfam; PF00036; ehand; 1.
DR      Pfam; PF01023; S_100; 1.
DR      ProDom; PD003407; CAP_S100; 1.
DR      ProSite; PS00018; EF_HAND; 1.
DR      ProSite; PS00303; S100_CAP; 1.
DR      PROSITE; PS00303; S100_CAP; 1.
SQ      SEQUENCE 2496 AA; 247588 MW; 4CE136CA6CE657DE CRC64;

Query Match      2.6%; Score 97.5; DB 11; Length 2496;
Best Local Similarity 21.5%; Pred. No. 1.2;
Matches 65; Conservative 30; Mismatches 145; Indels 63; Gaps 12;

Db      58 GEPGALAXGAKRPAARPGATRSKAPRWGLGRYT-CQTXQVSGELRE-DRLPLW 115
QY      723 GSTGGQTASTRQGS---GQGASGSGRCGASGQTSGGSGSTRYGEQSGSGRNSST 778
Db      116 XRAKARILIFNTDHEGASRSPFPGFAXXVAKVTGTGLMRPSXSDV-AFXS 174
QY      779 QSRGR-----STRSSTSR---FSGSGSGSGFSGRSGQGRSGRSGQGSFSG 826
Db      175 FDVGSYTHXEAFTKRWIVHPLIGXSWVTVVRQVAFLLXMCCHGNPAQYERNR-RX 233
QY      827 QTEGSGQHGS-----CC-GQSSGYQNEYGS 851
Db      234 RHLYVLGXGANGAKXXSVGLXNLSKSESRPXGTIRQRGASVGLGXPRLSPPAGR 293
QY      852 GHSASSGQGQSHYQSSSYGTHNSGSPSPSPAGHSGRSGSLG---QYSGPGQT 907
Db      294 PPSTXXXRAGGRV---RRAFPGSXACPSWETGRGRKXGPIARAPHVRAAEYXX 350
QY      908 SSTRQSGSGGQQAAGSGRSGVSGQTSGCGGQSTRYGEQSGS-RNSSTQGRSGRSTRS 966
Db      351 SST 353
QY      967 SST 969

RESULT 12
ID      Q84357      PRELIMINARY;      PRT;      542 AA.
AC      Q84357;
DT      01-AUG-1998 (TREMBlrel. 07, Created)
DT      01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      E2 protein.
GN      E2.
OS      Mastomys natalensis papillomavirus (Mnpv).
OC      Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC      Papillomavirus.
OX      NCBI_TaxID=10567;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=94120727; PubMed=8291235;
RA      Tan C.H., Tachezy R., Van Ranst M., Chan S.Y., Bernard H.U.;
RA      Burk R.D.;
RT      "The Mastomys natalensis papillomavirus: nucleotide sequence, genome
RT      organization, and phylogenetic relationship of a rodent papillomavirus
RT      involved in tumorigenesis of cutaneous epithelia."
RL      Virology 198:534-541(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      Chan S.;
RA      Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RL      EMBL; U01834; AAG7147.1; -.
DR      HSBP; P03122; ZBP.
DR      GO; GO:0042025; C:host cell nucleus; IEA.
DR      GO; GO:0003700; F:transcription factor activity; IEA.
DR      GO; GO:0006275; P:regulation of DNA replication; IEA.
DR      GO; GO:0006355; P:regulation of transcription; DNA-dependent; IEA.
DR      InterPro; IPR000427; E2_C.

```

DR InterPro: IPR001866; E2_N.
 DR InterPro: IPR009021; Viral_DNA_bd.
 DR Pfam: PF00511; E2_C; 1.
 DR Pfam: PF00508; E2_N; 1.
 DR ProDom: PD000672; E2_C; 1.
 DR ProDom: PD000678; E2_N; 1.
 SQ SEQUENCE 542 AA; 60991 MW; E7BB15DF005B26B CRC64;

Query Match 2.6%; Score 96.5; DB 12; Length 542;
 Best Local Similarity 25.0%; Pred. No. 0.24; 89; Indels 63; Gaps 11;
 Matches 56; Conservative 16; Mismatches 89; Indels 63; Gaps 11;

QY 160 GLM-RPSXHSVAFKSFVYSS-----YHKEAE---FTKRWIVHLLGXGSMVATVVR- 208
 DB 144 GMSRTTSHDINGITFNKSGDEYVYFKEBAKRYSLGTWEVHDLGTEHSLILPTSS 203
 QY 209 --QVSTFLXMCCHGNPAQYERNRXRLVYVLGKANGAKKXSVGLXNLSKESR-- 264
 DB 204 TPQTGFPP-----RQDPVRLHGN-----TTTGLPIPLRNSSNOI 237
 QY 265 -----PXGTRQRGASVGLGXPR-LSPPAGRPPTRXRXRAGRVPRRAF-- 312
 DB 238 LLEGSGDYDQARERRRYQG-PPTPRSLSPPIYRPPSYEERRRKRLRRQDGRV 296
 QY 313 --FGSXAECPSWET-----GRGRKGXPLARHAPHYRARE 347
 DB 297 KYAPSPYRTKPGETSDEDEGRGHEPRFRLPRGLRDGE 340

RESULT 13
 Q9RSH9 PRELIMINARY; PRT; 552 AA.

AC Q9RSH9; 09RSH9; 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter, ATP-binding protein.
 GN DR2145.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 CX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567286;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AB002048; AAF1688.1; -.
 DR PIR: F75311; F75311.
 DR TIGR: DR2145; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
 DR GO: GO:000165; F:nucleotide binding; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran_1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 552 AA; 58509 MW; 0688677C8C51C471 CRC64;

Query Match 2.6%; Score 96; DB 16; Length 552;
 Best Local Similarity 32.1%; Pred. No. 0.27;
 Matches 42; Conservative 7; Mismatches 44; Indels 38; Gaps 9;

QY 228 ERNRKRLVYVLGKANGAKKXSVGLXNLSKESRFXGTRIRORQASVG--LGXPXP 285
 DB 173 DRARRA-----AGHGRTPAVGSDATTTVPGR--GDYDRGALAGARLSHPLAR 223
 QY 286 LSP--AGRPPTRX--BAGRVPR--APPGSXAECPSWETGRGR 329
 DB 224 AAPGDAG-PRPDARAPARARIHWPERSGPRPRRAAPG-----C-----GSHGV 272
 QY 330 KGGXPLARHAP 340
 DB 273 GAGDPPARGSP 283

RESULT 14
 Q9HXB4 PRELIMINARY; PRT; 160 AA.

AC Q9HXB4; 09HXB4; 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopithecinae; Macaca.
 CX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TIGR=Frontal cortex;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TIGR=Frontal cortex;
 RX MEDLINE=21458551; PubMed=11574149;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
 RA Hirai M., Terao K., Suuki Y., Sugano S., Hashimoto K.;
 RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
 chromosomes.";
 RL Gene 275:31-37(2001).
 DR EMBL: AB097540; BAC41765.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 160 AA; 16645 MW; 21AABF7415CAFAC CRC64;

Query Match 2.5%; Score 95.5; DB 6; Length 160;
 Best Local Similarity 27.0%; Pred. No. 0.069;
 Matches 44; Conservative 11; Mismatches 43; Indels 65; Gaps 8;

QY 214 LIXMCCGNGNPAQYERNRXRLVYVLGXGA-----NGAKXSVGLXNLSKES 263
 DB 8 LILHL-----SAKQRERE-----IGVAGGARSSALQGGRGSGTARSIESPASRGS 58
 QY 264 RFXGTRQRGASVGLGXPR--RLSPAPARPPS-----TXXPARGRVPR----- 308
 DB 59 EAPGA--AGRGASVPGLPQCTWBPAPSPRPFRVRVCRVTRHGGSGVCAALPAFR 116
 QY 309 -----DAPGGSXAECPSWETGRGRKGGXP 334
 DB 117 SLAGSLCKPSGVGTAGBAAGGS-----RRSGRP 147

RESULT 15
 Q9XHX5 PRELIMINARY; PRT; 936 AA.

AC Q9XHX5; 09XHX5; 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative polypeptide.

GN OSJNB0049B20.18.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell R., Benito M.-I., Lin X., Mason T.M., Umayam L., Shea T.F.,
 RA Fujii C.Y., Shen M., Fraser C.M.,
 RT "Oryza sativa HAC OSJNB0049B20 genomic sequence."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007789; AAD38291.1; -.
 DR Gramene; OryzHX5; -.
 DR InterPro; IPR005162; Retrotrans_gag.
 DR Pfam; PF03732; Retrotrans_gag; 1.
 KM Polyprotein
 SQ SEQUENCE 936 AA; 102715 MW; 374D7D376F3FE7 CRC64;

Query Match 2.5%; Score 95; DB 10; Length 936;

Best Local Similarity 23.7%; Pred. No. 0.66;

Matches 81; Conservative 30; Mismatches 141; Indels 90; Gaps 19;

```

QY 1 IREHGEETKEVYN-KGEAPGAPVSPRGARGRRPCGPVKYHSDXXTDPR--RG 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 486 VRKRGHGTPPYEWQLRPDPPLRQGVGGRGRSIP---ATRWSSPSRLRG 540
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 GEPRGAL-AXGARPA-----ARRPATRSGXSARWGYLGRYTCQYXXQVSGEL 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 541 PPERGGLRATKRKRFTPLDGLYLERRRRAALR--AARYQQLRRY----- 584
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 REDRNLPMKRRAKRLIL---FSTNTDSESGASRSFPGFXKXKRYKVTG---ITGL 161
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 585 -----HORHVARSLCVDDVLRVQTRAGLSK-----LSPMEGPRYVIGV 626
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 162 WRPSXSHDYAFKSPDVGSYHKEAEFTKRWIVHPLIGXXSW-VXTVVRQVSFTLLMCC 220
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 627 PRP---GSVRLATGD-----GTLPNPNMNIHLRFYPLVAVVTTVALLV--- 672
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 221 HGNPAQYERNRRKRLVYV--LGXGANAKKXSVGLXNASKSESFPXGTIRQ----- 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 673 ---GVGEREPRTLEAVDDVGVDPDLRACAV-LGSERELFORRPGEVVRBAKFN 728
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 273 ---RGASVGLGXPRLSPPAGRPSTPSTXRXRAGRVPRRAP 311
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 729 GGRPPTAGVG-PPPACPTVSGAPDPQDSLGATAGR-PLSP 766
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: April 27, 2004, 10:55:16
 Job time : 43.7603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 1, 2004, 15:48:11; Search time 27 Seconds

(without alignments)
8921.733 Million cell updates/sec

Title: US-10-028-952A-4

Perfect score: 4230
Sequence: 1 gcacgagatccactgtcc.....agcttgcactcttcgcca 2333

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODE=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US10028952/raunat.01042004.154626.682/app.query.fasta_1.2503
-DB=Issued_Patents_AA -QWRT=fastan -SUPPFI=ra1 -MINMATCH=0.1 -LOOBL=0
-LOOEXT=0 -UNITS=Site -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.ccd
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=200000000
-USER=US10028952 @CGN 1 1 27 @raunat.01042004.154626.682 -NCFU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1189.5	28.1	355	4	US-08-630-915A-192
2	1033	24.4	538	4	Sequence 192, App
3	1033	24.4	538	4	Sequence 2, Appl
4	702.5	16.6	433	4	Sequence 18, Appl
5	324	7.7	58	4	Sequence 118, App
6	324	7.7	58	4	Sequence 214, App
7	277	6.5	486	2	Sequence 3, Appl
8	261.5	6.2	486	2	Sequence 2, Appl
9	261.5	6.2	486	2	Sequence 26, Appl
10	237	5.6	726	4	Sequence 20675, A
11	236.5	5.6	1418	4	Sequence 32367, A
12	231	5.5	546	2	Sequence 4, Appl

13	231	5.5	546	4	US-08-630-915A-14	Sequence 14, Appl
14	228.5	5.4	798	4	US-09-252-991A-23774	Sequence 23774, A
15	224.5	5.3	663	4	US-09-252-991A-30843	Sequence 30843, A
16	223.5	5.3	937	4	US-09-252-991A-19446	Sequence 19446, A
17	220.5	5.2	957	4	US-09-252-991A-20408	Sequence 20408, A
18	220	5.2	1706	4	US-09-252-991A-31760	Sequence 31760, A
19	219.5	5.2	2294	4	US-09-252-991A-17231	Sequence 17231, A
20	219	5.2	686	4	US-09-252-991A-20509	Sequence 20509, A
21	214	5.1	863	4	US-09-252-991A-26099	Sequence 26099, A
22	213	5.0	1002	4	US-09-252-991A-27980	Sequence 27980, A
23	212	5.0	1008	4	US-09-252-991A-29419	Sequence 29419, A
24	209.5	4.9	720	4	US-09-252-991A-31915	Sequence 31915, A
25	208.5	4.9	783	4	US-09-252-991A-28327	Sequence 28327, A
26	206	4.8	783	4	US-09-252-991A-18035	Sequence 18035, A
27	204.5	4.8	57	4	US-08-630-915A-116	Sequence 116, App
28	204	4.8	594	4	US-09-252-991A-32578	Sequence 32578, A
29	203.5	4.8	555	4	US-09-252-991A-28848	Sequence 28848, A
30	203.5	4.8	891	4	US-09-252-991A-30624	Sequence 30624, A
31	203.5	4.8	1197	4	US-09-252-991A-30833	Sequence 30833, A
32	203	4.8	695	4	US-09-252-991A-22910	Sequence 22910, A
33	202	4.8	654	4	US-09-252-991A-25801	Sequence 25801, A
34	201.5	4.8	1411	4	US-09-252-991A-18008	Sequence 18008, A
35	201	4.8	433	4	US-09-252-991A-27024	Sequence 27024, A
36	199.5	4.7	537	4	US-09-252-991A-20509	Sequence 20509, A
37	199.5	4.7	686	4	US-09-252-991A-20509	Sequence 20509, A
38	199.5	4.7	802	4	US-09-252-991A-20509	Sequence 17817, A
39	199	4.7	472	4	US-09-252-991A-17817	Sequence 30690, A
40	199	4.7	588	4	US-09-252-991A-30690	Sequence 29419, A
41	199	4.7	1008	4	US-09-252-991A-29419	Sequence 37, Appl
42	198.5	4.7	53	2	US-08-942-423-37	Sequence 18557, A
43	198	4.7	679	4	US-09-252-991A-18857	Sequence 31954, A
44	198	4.7	703	4	US-09-252-991A-31954	Sequence 21881, A
45	198	4.7	720	4	US-09-252-991A-21881	

ALIGNMENTS

RESULT 1
US-08-630-915A-192
; Sequence 192, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFEMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWKES, Dana M.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 192:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-630-915A-192

Alignment Scores:

Pred. No.:	1.98e-82	Length:	355
Score:	1189.50	Matches:	251
Percent Similarity:	73.28%	Conservative:	15
Best Local Similarity:	69.15%	Mismatches:	63
Query Match:	28.12%	Indels:	34
		Gaps:	6

US-10-028-952A-4 (1-2333) x US-08-630-915A-192 (1-355)

```

QY 724 CTCAGTACGAGAGGAAACCGCAGCTTCAGACATTTGGTGTATG-----TCGTTG 771
DB 12 LeuProGlyArgGlyThrProGlyProSerGlyLeuGlyValProGluAspGlnCysArg 31
QY 772 GCTAGAGGACCAATGGGGCGAAGCTACATCTGTGGGATTATGACTGAAGCCCTTAAGT 831
DB 32 ValArgAspLeuLysGlyTyrLeuAspSerPheThrAlaLysAlaGluLysGluGlu 51
QY 832 CAGATCCCGCCGCGGCGAAGCATACGCGACGCCCGCGGAGCTCCGTTGGCTCGGAT 891
DB 52 AsnArgArgLeuGluGluLysArgTyrAlaGluLysAlaGlnArgGlnLeuGluGln 71
QY 892 AGCCGCTCCCGCCGCTGTCGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
DB 72 ArgArgGlyLysArgGluLeu-----ArgGluAlaAlaArgArgGluGlnArgTyrGlnGlu 89
QY 952 GCGGAGAGGCG--GCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 979
DB 90 GlnGly-GlyGluAlaSerProGlnSerArgThrTyrGluGlnGlnGlnValAlaIse 109
QY 980 -CGGAGCCGGGTCGCGCGGAGTGCCTCGTCCGTAAGAGGGGCGCGCGGAGAA 1038
DB 109 ArgAsnArgAsnGlnGlnGlnSerAlaValHisProArgGluLeu-PheLysGlnLysG 129
QY 1039 GAGCGCCCGCCCTGCGCCGTCACGACCGACGCTTCGTGGGAAACCTGGCGCTAAACA 1098
DB 129 LysArg-----AlaMetSer-----ThrT 135
QY 1099 CCTCATCTCCAGTCTCTAGCTTCGAGAGCTGAGAGAGCCCTTCCTGAGAGAGAGCTCA 1158
DB 135 HisSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnLysGlnLeuT 155
QY 1159 CCCAACACAGAGACCCACTTTGGCAGAGAGCAAGCTGCTGCATCTCAAGGCCCGAGG 1218
DB 155 HisGlnProGlnThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProArgLaa 175
QY 1219 ATCTCCCTGCTAGAGAGCGCGGCGCCAGACATCTCCCTCCATGCTGTGTGAGGCAAGAAG 1278
DB 175 sPheLysProAlaGlnGlnProAlaProSerThrProProCysLeuValGlnLysGln 195
QY 1279 AGGCTGTGTATAGAGAACCTCCAGAGAGAGAGACCTTCTACAGAGAGAGCCCGCAGTGTGC 1338
DB 195 LysAlaValLysGlnGlnProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 215
QY 1339 AGCAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1398
DB 215 LysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 235
QY 1399 GCGAAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1458
DB 235 LysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 255

```

```

QY 1459 CCTTGACCCCGAGAACTCATCAAGGAGCATTCGAGTGTATGACGAAGAGCTGTGCGCTG 1518
DB 255 erPheAspProGluAsnLeuIleThrGlyIleGluValIleAspGlnGlyTyrThrArg 275
QY 1519 GCTATGGGCGCGATGAGCCATTTGGCATGTTCCTCCCACTACGTGAGACTCATTTAGT 1578
DB 275 LysTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIleAsp 294
QY 1579 GAGGCTGAGGACACATCTTGGCTCCCTCCCTGACAGATGAGCTTCTTATGCTGAAGA 1638
DB 295 GlnAlaGlnGlnGlnHisSerCysProSerProLeuArgHisGlyPheLeuIleAlaGlyArg 314
QY 1639 GAGGCTGAGGAGTTCATTCAGACTCTTCCAGAAATAGACCCCGCAGTGAAGATGAG 1698
DB 315 GlyGlyLeuGlyValAspIleGlnHisSerSerArgAsnArgThrProSerGluAspGln 334
QY 1699 GCGTCAAGGCTCCCTCCCGCTTGGAGAGCTGAGCTGACCCCAATGACGAATGAGCC 1758
DB 335 AlaSerGlyLeuProProAlaTyrPdnThrGlnProValThrProAsnAlaIleMetAla 354
QY 1759 TGG 1761
DB 355 TTP 355

RESULT 2
US-08-994-076-2
; Sequence 2, Application US/08994076
; Patent No. 6500937
; GENERAL INFORMATION:
; APPLICANT: Evid. Jf., Paul R.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE
; TITLE OF INVENTION: OF MAMMASTATIN AND METHODS OF USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6500937 West Center, 90 South 7th Street
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,076
; FILING DATE: 19-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,315
; FILING DATE: 03-OCT-1996
; APPLICATION NUMBER: PCT/US97/16026
; FILING DATE: 03-OCT-1997
; APPLICATION NUMBER: 08/943,828
; FILING DATE: 03-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ketelberger, Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 4273.1US11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 538 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-994-076-2

```

Alignment Scores:

Pred. No.: 1,74e-70 Length: 538
 Score: 1033.00 Matches: 212
 Percent Similarity: 95.52% Conservative: 1
 Best Local Similarity: 95.07% Mismatches: 7
 Query Match: 24.42% Indels: 4
 Gaps: 0

US-10-028-952a-4 (1-2333) x US-08-994-076-2 (1-538)

QY 417 TTGATCTTGATTTTCAGTACGAATACAGACCGTGAAGCGGGCCCTCAGATCCTTCTGA 476
 Db 1 MetIleuLeuIlePheSerThrAsnThrAspArgGluSerGlyAlaSer-Asp-LeuLeuT 20
 QY 477 CCTTTGGGTTTAAAGCAGAGGTGTCAAGAAAGTTACACAGAGGATTAAGCTTGTGG 536
 Db 20 hrPheTrpValLeuSerArgArgCysGlnYsSerYrHsArgAspAsnTrpLeuValA 40
 QY 537 CGGCAAGCGTTTCATAGCAGCGTGCCTTTTGATCCTTCGATGTCGGCTCTTCATCAT 536
 Db 40 laaIalysArgSerIysArgArgArgPheLeuIleLeuArgCysArgLeuPheLeuSerI 60
 QY 597 TGTGAAGCAGAAATTCACCAAGCGTTGATTTGATCCACTAATAGGAACGTGAGCTGG 656
 Db 60 eugIysSerArgIleHsGlnAlaLeuAspCysSerProThrAsnArgGluArgGluLeuG 80
 QY 657 GATTAGACCGTGCAGACAGGTTAGTTTACCTACTGATGATGTGTGTGTCATGCT 716
 Db 80 lylLeuAspArgArgGlnThrGly-LeuPheThrLeuLeuMetCysCysGlySgIy 99
 QY 717 AATCTGCTCAGTACGAGAGGAAACCGAGGTTTCAGACATTTGTGTATGCTTGGCTGA 776
 Db 100 AsnProAlaGlnIlyGluArgAsnArgArgPheArgHisLeuValIlyrValLeuGlyTrp 119
 QY 777 GAGGCCAATGGGGCGCAAGCTACCACTTGTGGATTAATACAGACGGCTCTTAAGTCAA 836
 Db 120 GlyAlaAlaGlnGlyAlaLeuProSerValGlyLeuLeuAlaSerIysSerGlu 139
 QY 837 TCCCGCCGAGGCGGAGCATACGAGCGCGCGAGCGCTGGCTGGCTGGATACCG 896
 Db 140 SerArgProGlyGlyThrIleArgGlnArgArgGlyAlaSerValGlyLeuGlyTrpPro 159
 QY 897 GTCCCGCGCTTCTCCCGCGCGCGCGCGCG-CCCCCGCTTCACAGCGCGCGCGCGCGG 955
 Db 160 ValProArgLeuSerProProAlaGlyAlaProProProProArgAlaProArgAlaArg 179
 QY 956 GAGGCGCGTGCCTCCCGCGCGCGCGGAGCGGGGTCCGCTGCGGAGTCCCTTCGTCCT 1015
 Db 180 GluGlyAlaCysProAlaAlaArgArgAspArgGlyProValArgSerAlaLeuArgPro 199
 QY 1016 GGGAAACGGGGCGCGCGCGGAAAGCGCGCGCGCTTCGCTGACGACGACGACGCTTC 1075
 Db 200 GlyIysArgGlyAlaAlaGlyIysAlaAlaProSerProValThrHisArgThrPhe 219
 QY 1076 GTG 1078
 Db 220 Val 220

RESULT 3

US-09-643-476-2
 ; Sequence 2, Application US/09643476
 ; Patent No. 6599495

GENERAL INFORMATION:

APPLICANT: Eryn, Jr., Paul R.
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE
 OF MAMMASTATTIN AND METHODS OF USE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant & Gould P.C.

STREET: P.O. Box 2903

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402-0903

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,476

FILING DATE: 22-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,315

FILING DATE: 03-OCT-1996

APPLICATION NUMBER: PCT/US97/18026

FILING DATE: 03-OCT-1997

APPLICATION NUMBER: 08/943,828

FILING DATE: 03-OCT-1997

APPLICATION NUMBER: 08/994,076

FILING DATE: 19-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Weaver, Karie G.

REGISTRATION NUMBER: 43,245

REFERENCE/DOCKET NUMBER: 4273, IUSC2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 538 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-643-476-2

Alignment Scores:
 Pred. No.: 1,74e-70 Length: 538
 Score: 1033.00 Matches: 212
 Percent Similarity: 95.52% Conservative: 1
 Best Local Similarity: 95.07% Mismatches: 7
 Query Match: 24.42% Indels: 4
 Gaps: 0

US-10-028-952a-4 (1-2333) x US-09-643-476-2 (1-538)

QY 417 TTGATCTTGATTTTCAGTACGAATACAGACCGTGAAGCGGGCCCTCAGATCCTTCTGA 476
 Db 1 MetIleuLeuIlePheSerThrAsnThrAspArgGluSerGlyAlaSer-Asp-LeuLeuT 20
 QY 477 CCTTTGGGTTTAAAGCAGAGGTGTCAAGAAAGTTACACAGAGGATTAAGCTTGTGG 536
 Db 20 hrPheTrpValLeuSerArgArgCysGlnYsSerYrHsArgAspAsnTrpLeuValA 40
 QY 537 CGGCAAGCGTTTCATAGCAGCGTGCCTTTTGATCCTTCGATGTCGGCTCTTCATCAT 536
 Db 40 laaIalysArgSerIysArgArgArgPheLeuIleLeuArgCysArgLeuPheLeuSerI 60
 QY 597 TGTGAAGCAGAAATTCACCAAGCGTTGATTTGATCCACTAATAGGAACGTGAGCTGG 656
 Db 60 eugIysSerArgIleHsGlnAlaLeuAspCysSerProThrAsnArgGluArgGluLeuG 80
 QY 657 GATTAGACCGTGCAGACAGGTTAGTTTACCCCTACTGATGATGTGTGTGTCATGCT 716
 Db 80 lylLeuAspArgArgGlnThrGly-LeuPheThrLeuLeuMetCysCysGlySgIy 99
 QY 717 AATCTGCTCAGTACGAGAGGAAACCGAGGTTTCAGACATTTGTGTATGCTTGGCTGA 776
 Db 100 AsnProAlaGlnIlyGluArgAsnArgArgPheArgHisLeuValIlyrValLeuGlyTrp 119

QY 777 GGAGCCATAGGGGGAAGTACCATCTGTGGATTATGATGAACGCTCTTAAGTGA 836
 Db 120 G1YAlaHsnG1YAlaYsLeuProSerValG1YLeuLeuHsnAlaSerLysSerGlu 139
 QY 837 TCCCGCCGAGGCGGAAGATACGCGACCGCGGAGCCTCGGTGGCCCTCGATAGCGG 896
 Db 140 SerArgProG1Yg1YThrL1aArgG1nArgG1YAlaSerValG1YLeuG1YTrpPro 159
 QY 897 GTCCCGCCCTGTGCTCCCGCCGCGCGCGCG-CCCCCGCTCCACGCGCGCGCGCGCGG 955
 Db 160 ValProArgLeuSerProProAlaG1YAlaProProProProAlaProAlaArg 179
 QY 956 GAGGCGCGCTGCTCCCGCGCGCGCGCGGACCGGAGTCCGCGCGATGCTTCGCT 1015
 Db 180 G1nG1YAlaCySProAlaAlaArgArgSPargG1YProValArgSerAlaLeuArgPro 199
 QY 1016 GGGAAACGGGCGCGCGCGGAAAGCGCGCGCGCGCTCGCGCTCAGCAGCAGCGCGCTTC 1075
 Db 200 G1YysArgG1YAlaAlaG1YlysAlaAlaAlaProSerProValThrHisArgThrPhe 219
 QY 1076 GTG 1078
 Db 220 Val 220
 RESULT 4
 US-08-630-915A-18
 / Sequence 18, Application US/08630915A
 / Patent No. 6309820
 / GENERAL INFORMATION:
 / APPLICANT: SPARKS, Andrew B.
 / APPLICANT: HOFFMAN, No. 6309820h
 / APPLICANT: KAY, Brian K.
 / APPLICANT: FOWLES, Dana M.
 / APPLICANT: MCCONNELL, Stephen J.
 / TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 / TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 / NUMBER OF SEQUENCES: 227
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Pennie & Edmonds LLP
 / STREET: 1155 Avenue of the Americas
 / CITY: New York
 / STATE: New York
 / COUNTRY: USA
 / ZIP: 10036-2711
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent Release #1.0, Version #1.30
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/630,915A
 / FILING DATE: 03-APR-1996
 / CLASSIFICATION: 536
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Mastro, S. Leslie
 / REGISTRATION NUMBER: 18,872
 / REFERENCE/DOCKET NUMBER: 1101-174
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (212) 790-9090
 / TELEFAX: (212) 869-8664/9741
 / TELEX: 66141 PENNIE
 / INFORMATION FOR SEQ ID NO: 18:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 433 amino acids
 / TYPE: amino acid
 / STRANDEDNESS:
 / TOPOLOGY: unknown
 / MOLECULE TYPE: peptide
 / US-08-630-915A-18
 Alignment Scores:
 Pred. No.: 2.15e-45 Length: 433

Score: 702.50 Matches: 146
 Percent Similarity: 72.65% Conservative: 16
 Best Local Similarity: 65.47% Mismatches: 35
 Query Match: 16.61% Indels: 26
 DB: 4 Gaps: 5
 US-10-028-952a-4 (1-2333) x US-08-630-915A-18 (1-433)
 QY 971 GCGCGCGCGCGGAC-----CGGGGTCCGGTCCGAGTGCCTTCCTCGGAA 1021
 Db 217 AlaAlaArgArgG1nG1nArgG1nG1nG1nG1nHisArgSerAlaG1YAlaProSerArg 236
 QY 1022 CGGGGCGGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1081
 Db 237 ThrG1Y-G1nProG1nG1nG1n-----AlaValSerArgThrArgG1nG1nG1nG1n 253
 QY 1082 AAC-----TGCG 1090
 Db 253 uSerAlaG1nG1nAlaProHisProArgG1nLeuPheYsg1nYsg1nArgAlaMe 273
 QY 1091 CTAAACCACTCCATCTCCAGTCTCAGCTGGAGCTGAGAGCCCTTCCTGAGAA 1150
 Db 273 tSerThrThrSerValImrSerSerGlnProG1YysLeuArgSerProPheLeuG1n 293
 QY 1151 GCAGCTCACCAACAGAGAGAGCCACTTGGAGAGAGCGAGCTGCTGCATCTCAGGCC 1210
 Db 293 sGlnLeuThrGlnProG1nThrSerThrG1YArgG1nProThrAlaProValSerArgPr 313
 QY 1211 CAGGCGAGTCTCCCTGCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1270
 Db 313 oAlaAlaG1YVal---CySG1nG1nProAlaProSerThrLeuSerSer---AlaG1nTh 331
 QY 1271 AGAAGAGAGGCTGTGTATAGAGAACTTCAGAGAGAGAGACCTTTCAGAGAGAGAG 1330
 Db 331 rG1nG1nG1nProThrThrG1YValProProG1nG1nSerThrLeuYrG1nG1nProPr 351
 QY 1331 ACTGTGACAGAGAGAGAGTGTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1390
 Db 351 oLeuValG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1nG1n 371
 QY 1391 GCTCAGTGGGCAAGGCTGTGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1450
 Db 371 yPheSerG1YnG1nG1YLeuCyAlaArgAlaLeuYrAspYrG1nAlaAlaAspAsp 391
 QY 1451 AGAGATCTCTTTGAGACCCCGAGAACTCATACAGGAGAGAGAGAGAGAGAGAGAG 1510
 Db 391 rG1nLeuSerPheAspProG1nAsnLeuLeuThrG1YLeuValAlaAspG1nG1Yr 411
 QY 1511 GTGCGGTGCTATGCGCGAGATGAGCATTTTGGAGATGTCCTGCGCACTAGAGAGCT 1570
 Db 411 pTrpArgG1YrG1YProAspG1YHisPheG1YMetPheProAlaAsnYrValG1nG1n 431
 QY 1571 CATTGAG 1577
 Db 431 uileGlu 433
 RESULT 5
 US-08-630-915A-118
 / Sequence 118, Application US/08630915A
 / Patent No. 6309820
 / GENERAL INFORMATION:
 / APPLICANT: SPARKS, Andrew B.
 / APPLICANT: HOFFMAN, No. 6309820h
 / APPLICANT: KAY, Brian K.
 / APPLICANT: FOWLES, Dana M.
 / APPLICANT: MCCONNELL, Stephen J.
 / TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 / TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 / NUMBER OF SEQUENCES: 227
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Pennie & Edmonds LLP
 / STREET: 1155 Avenue of the Americas

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIB
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-118

Alignment Scores:
Pred. No.: 7,11e-17 Length: 58
Score: 324.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.66% Indels: 0
Gaps: 0

US-10-028-952a-4 (1-2333) x US-08-630-915A-118 (1-58)

QY 1401 CAAGGGCTGTGTCGCGCCCTGTACGACTACAGGCGAGCGAGACAGAGATCTCC 1460
Db 1 GIndYLeucYsAlaTgAlaLeuTYrAspTYrGlnAlaAlaAspAspThrGlnLLeSer 20

QY 1461 TTGACCCCGAGAACCTCATCAAGGCGATCGAGTGATGACGAGAGCGCTGTGGCGTGGC 1520
Db 21 PheAspProGlnAsnLeuIleThrGlyIleGluValIleAspGluGlyTPrTParGly 40

QY 1521 TATGGCCCGATGGCCATTTTGGCATGTTCCTGCCCAACTAGCTGAGCTCATT 1574
Db 41 TYrGlyProAspGlyHisPheGlyMetPheProAlaAsnTYrValGluLeuIle 58

RESULT 6
US-08-630-915A-214
Sequence 214, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penite & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIB
INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-214

Alignment Scores:
Pred. No.: 7,11e-17 Length: 58
Score: 324.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.66% Indels: 0
Gaps: 0

US-10-028-952a-4 (1-2333) x US-08-630-915A-214 (1-58)

QY 1401 CAAGGGCTGTGTCGCGCCCTGTACGACTACAGGCGAGCGAGACAGAGATCTCC 1460
Db 1 GIndYLeucYsAlaTgAlaLeuTYrAspTYrGlnAlaAlaAspAspThrGlnLLeSer 20

QY 1461 TTGACCCCGAGAACCTCATCAAGGCGATCGAGTGATGACGAGAGCGCTGTGGCGTGGC 1520
Db 21 PheAspProGlnAsnLeuIleThrGlyIleGluValIleAspGluGlyTPrTParGly 40

QY 1521 TATGGCCCGATGGCCATTTTGGCATGTTCCTGCCCAACTAGCTGAGCTCATT 1574
Db 41 TYrGlyProAspGlyHisPheGlyMetPheProAlaAsnTYrValGluLeuIle 58

RESULT 7
US-08-942-423-3
Sequence 3, Application US/08942423
Patent No. 5891673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro
APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Lock Binding Protein
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435

Fri Apr 2 07:35:03 2004

us-10-028-952a-4.ra1

Page 6

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEITICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: LCK BINDING PROTEIN

US-08-942-423-3

Alignment Scores:
Pred. No.: 4,81e-13 Length: 486
Score: 277.00 Matches: 87
Percent Similarity: 42.97% Conservative: 23
Best Local Similarity: 33.98% Mismatches: 85
Query Match: 6.55% Indels: 61
DB: 2 Gaps: 10

US-10-028-952A-4 (1-2333) x US-08-942-423-3 (1-486)

QY 938 CGCGCCCGCGCGCGCGGAGGCGCGGCCCCCGCGCGCGG-----GACCGGGGT 991
Db 250 Arglysharlguglulgululylalaglnlvalalavargaglnglnglularglys 269
QY 992 CCGGTG-----CGAGTGCCTTCCTGCTGGAAAGCGGCGCGCGGAAAGCGGCC 1045
Db 270 AlalvalThrllysArGysSerArglulAlaProglInProvalIlleAlaMetglulnProAla 289
QY 1046 GCCCCCTGCGCCGTCACGACGACCGACGACGTCCTGGGGAACCTGGCGCTAAACCACTCCAT 1105
Db 290 ValProAlaProLeuProLys----- 297
QY 1106 CTCACGTCTTCAAGCTTGCGAAGCTGAGAGGCCCTTCTCTGAGAAAGCAAGCTCACCCAAAC 1165
Db 298 IlleSerSerGlulAlaTrpProProValglYThrProProSerSergluSerGlulProVal 317
QY 1166 AGAGACCACATTGGCAGAGAGCCAGCTGTCGCACTTCAAGAGCCCGAGGCGAGATCTCC 1225
Db 318 Arg-Thr-----SerArglulHleProvalProleuLeuProIleArglnThrLeuPr 335
QY 1226 TGCT-----GAGAGCGCGCGCGCCAGACACTCTTCAATGTG-----GTGCGAGGC 1270
Db 335 OGluAspasnGlulnPro-----ProAlaLeuProProArgThrLeuGlulnLeuGlnVa 354
QY 1271 AGAAGAGAGAGGCTGTGTATGAGGAACCTCCAGAGAGAGAGACCTTCTACGAGAGAGCCCC 1330
Db 354 IGlulnGlulnProValTYrglulAlaGlulnProglulnPro-----GluProglulnPr 372
QY 1331 ACTGTGCAAGCAAGAGTGTGCTGCTGAGACACTTGAACCAACCACTTCAG----- 1382
Db 372 OGluProglulnAspTYrglulnAspValglulnUmetAspArgHlsGlulnGlulnAspGl 392
QY 1382 ----- 1382
Db 392 uProglulnAspTYrglulnValleuGlulnProglulnAspSerSerPheSerSerAla 412
QY 1383 -----GGCAGGCGCTCACTGGGCAAGGCGCTCG 1411
Db 412 uAlaGlySerSerGlyCySPeAlaGlyAlaGlyAlaGlyAlaValAlaLeuGlyIleSe 432

QY	1412	TGCCCCGCGCCCTGATGACATCACCGAGCGCCGACGACACACGAGATCTTCCTTGAACCCCGA	14711
Db	432	raIaValaIaIaenUryrAspTyrgInglYglUglYseArSpglUluuSeArPheAspProAs	452
QY	1472	GAACCTCATCACGGGCATCGAGGTGATCGACGAAAGGCTGGTGGCGCTATGGAGCCCGGA	15311
Db	452	PaepValIlethrAspIleGluuSeValaSpglUglYTrItrPrgrglY--ArgCysH	471
QY	1532	TGGCGCATTTGGCATGTTCCTGGCCACTGACGTGAGACTCATTTAG	1577
Db	471	sgIYhIspheGlyuPheProhIaAsnTyValIyLeuIeuIdU	486
RESULT 8			
US-08-942-423-2			
Sequence No. Application US/08942423			
Patent No. 5691673			
GENERAL INFORMATION:			
APPLICANT: Hashimoto, Yasuhiro			
APPLICANT: Takemoto, Yoshihiro			
TITLE OF INVENTION: Lock Binding Protein			
NUMBER OF SEQUENCES: 68			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Syntex (U.S.A.) Inc.			
STREET: 3401 Hillview Ave.			
CITY: Palo Alto			
STATE: California			
COUNTRY: U.S.A.			
ZIP: 94303			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/942,423			
FILING DATE: 01-OCT-1997			
CLASSIFICATION: 435			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/362,715			
FILING DATE: 23-DEC-1994			
ATTORNEY/AGENT INFORMATION:			
NAME: Peries, Rohan			
REGISTRATION NUMBER: 35,752			
REFERENCE/DOCKET NUMBER: 28260			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (415) 852-1698			
TELEFAX: (415) 496-3529			
INFORMATION FOR SEQ ID NO: 2:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 486 amino acids			
TYPE: amino acid			
STRANDEDNESS: not relevant			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
HYPOTHETICAL: NO			
ANTI-SENSE: NO			
FRAGMENT TYPE: N-terminal			
ORIGINAL SOURCE:			
ORGANISM: HSI			
US-08-942-423-2			
Alignment Scores:			
Pred. No.: 7,256-12 Length: 486			
Score: 261.50 Matches: 92			
Percent Similarity: 38.26% Conservative: 22			
Basic Local Similarity: 30.87% Mismatches: 107			
Query Match: 6.18% Indels: 77			
DB: 2 Gaps: 12			
US-10-028-952A-4 (1-2333) x US-08-942-423-2 (1-486)			
QY 777 GGAGCAATGGG-----GCGAAGCTACCATCTGTGGATATATGACTGACGCGCTCTTAAG 830			

330

```

Db      234 GLYAlaArgGlyLeuYsAlaIlePhedIuSerLeuAlaGluGluYsArgGlySargGlu 253
QY      831 TCAGAAATCCCGCCCAAGCGGGAACGATACGAGCAGCGC-----CGCGAGCCTCGATTGGC 884
Db      254 GluGluGluYsAlaGlnGlnMetAlaArgGlnGlnGlnGluArgGlySAlaValLys 273
QY      885 CTCGGATAGCCGGTCCCGCCCTGTCGCCCGGCGGCGCGCGCC-----CCCTCCAC 938
Db      274 MetSerArgGluValGlnGlnProSerMetProValGluGluProAlaAlaProAlaGln 293
QY      939 GCGCCCGCGCGCGCGGAGGAGCGCGTCCCGCGCGCGCGGAGACCGGGTCCGGTGC 998
Db      294 LeuProLysLysIleSerSerGluValTrpProPro----- 305
QY      999 GAGTGGCCCTTCCTCTGGGAAACGGGGCGCGCGGCGGAAAGCGCGCGCCCTCGCCCG 1058
Db      306 -----AlaGluSerHisLeuProProGluSerGlnPro 316
QY      1059 TCACGACCGGACGTTCTGCGGGAACCTGGCGCTAAACACCTCCATCTCCAGCTCTAG 1118
Db      317 ValArgSerArgArgGluTyr-----ProVal 325
QY      1119 CCTGCG---AAGCTGAGAGCCCTTCTCTGAGAGAGAGCTCACCCACAGAGACCCAC 1175
Db      326 ProSerLeuProThrArgGlnSerProLeuGlnSerHisLeuGluAspAsnGlu----- 343
QY      1176 TTGGCAGAGAGCGAGCTGCTCCATCTCAAG---CCGAGGAGATCTCTGCTGAG 1232
Db      344 -----GluProProAlaLeuProProAlaGlnThrProGluGluValValGlu 360
QY      1233 GAGCGCGCGCGCGAGCTCTCTCATGCTGTGCGAGGAGAGAGAGAGCTGTGATGAG 1292
Db      361 GluProValTyrGluAlaAlaProGluLeu---GluProGluProGluProAspTyrGlu 379
QY      1293 GAACCTCCAGAGAGAGAGAGAGCTTCTAGAGAG----- 1325
Db      380 ProGluProGluThrGluProAspTyrGluAspValGluGluLeuAspArgGluAspGlu 399
QY      1326 -----CCCCCATGCTGCGAG 1340
Db      400 AspAlaGluGlyAspTyrGluAspValLeuGluProGluAspThrProSerLeuSerTyr 419
QY      1341 CAGCAGAGGCTGCTGCTGAGACATTTGACACACCATTTACAGGCGGCGCTCAGTGG 1400
Db      420 GlnAlaGlyProSerAla-----GlyAlaGlyGly 429
QY      1401 CAAGGAGCTGTGCGCGGCTGCTGATCACTACAGGAGCGAGCGAGACAGAGATCTCC 1460
Db      430 AlaGlyIleSerAlaIleAlaLeuTyrAspTyrGlnGlyGluGlySerAspGluLeuSer 449
QY      1461 TTGGACCCGAGAGCTCATCAAGGAGCTGAGGAGCTGAGGAGGAGGAGGAGGAGTGGC 1520
Db      450 PheAspProAspAspIleIleThrAspIleGluMetValAspGluGlyTrpTrpArgGly 469
QY      1521 TAGGCGCGGATGCGCATTTGGCATGTTCCTGCGCACTACGTTGAGAGCTATT 1574
Db      470 Gln---CysArgGlyHisPheGlyLeuPheProAlaAsnTyrValLysLeuLeu 486

```

```

CORRESPONDENCE ADDRESS:
ADDRESSER: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEO ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-26
Alignment Scores:
Pred. No.: 7,25e-12 Length: 486
Score: 261.50 Matches: 92
Percent Similarity: 38.26% Conservative: 22
Best Local Similarity: 30.87% Mismatches: 107
Query Match: 6.18% Indels: 77
Gaps: 4 Gaps: 12
US-10-028-952a-4 (1-2333) x US-08-630-915A-26 (1-486)
QY      777 GGAAGCAATGG-----GCGAAGCTACCATCTGTGAGTTATGACTGAACGCTCTAAG 830
Db      234 GLYAlaArgGlyLeuYsAlaIlePhedIuSerLeuAlaGluGluYsArgGlySargGlu 253
QY      831 TCAGAAATCCCGCCCAAGCGGGAACGATACGAGCAGCGC-----CGCGAGCCTCGATTGGC 884
Db      254 GluGluGluYsAlaGlnGlnMetAlaArgGlnGlnGlnGluArgGlySAlaValLys 273
QY      885 CTCGGATAGCCGGTCCCGCCCTGTCGCCCGGCGGCGCGCGCC-----CCCTCCAC 938
Db      274 MetSerArgGluValGlnGlnProSerMetProValGluGluProAlaAlaProAlaGln 293
QY      939 GCGCCCGCGCGCGCGGAGGAGCGCGTCCCGCGCGCGCGGAGACCGGGTCCGGTGC 998
Db      294 LeuProLysLysIleSerSerGluValTrpProPro----- 305
QY      999 GAGTGGCCCTTCCTCTGGGAAACGGGGCGCGCGGCGGAAAGCGCGCGCCCTCGCCCG 1058
Db      306 -----AlaGluSerHisLeuProProGluSerGlnPro 316
QY      1059 TCACGACCGGACGTTCTGCGGGAACCTGGCGCTAAACACCTCCATCTCCAGCTCTAG 1118
Db      317 ValArgSerArgArgGluTyr-----ProVal 325
QY      1119 CCTGCG---AAGCTGAGAGCCCTTCTCTGAGAGAGAGCTCACCCACAGAGACCCAC 1175
Db      326 ProSerLeuProThrArgGlnSerProLeuGlnSerHisLeuGluAspAsnGlu----- 343
QY      1176 TTGGCAGAGAGCGAGCTGCTCCATCTCAAG---CCGAGGAGATCTCTGCTGAG 1232

```

```

Db 344 -----GluProProlAlaLeuProProArgThrProGluGluValGlu 360
QY 1233 GAGCCGGCGCCGACACTCTCCATGTCGTGCAGCAGAGAGAGGCTGTATGAG 1292
Db 361 GluProValTyrGluAlaAlaProGluLeu---GluProGluProGluProAspTyrGlu 379
QY 1293 GAACCTCCAGAGCAGAGACCTTCTAGCAGAG- 1325
Db 380 ProGluProGluThrGluProAspTyrGluAspValGluGluLeuAspArgGluAspGlu 399
QY 1326 -----CCCCCACTGTCGAG 1340
Db 400 AspAlaGluGluAspTyrGluAspValLeuGluProGluAspThrProSerLeuSerTyr 419
QY 1341 CAGCAGAGTGTGCTGCTGAGCATTGACCACTTACAGGCCGAGGGCTGCTAGTGG 1400
Db 420 GluAlaGlyProSerAla-----GlyAlaGlyGly 429
QY 1401 CAAAGGCTGTGTGCGCCGCTGTAGACTACGAGCAGCAGCAGCAGAGATCTCC 1460
Db 430 AlaGlyLeuSerAlaLeuAlaLeuTyrAspTyrGluGluGlySerAspGluLeuSer 449
QY 1461 TTGAACCCGAGAACCTTCATACAGGCGATCGAGTGTATCGACGAGGCTGTGCTGCTG 1520
Db 450 PheAspProAspAspIleIleThrAspIleGluMetValAspGluGlyTyrTrpArgGly 469
QY 1521 TATGGCCCGATGGCATTTGGCATGTCCTGCCCACTACGAGTGCATTT 1574
Db 470 Glu---CysArgGlyHisPheGlyLeuPheProAlaSerTyrValGlyLeuLeu 486

RESULT 10
US-09-252-991A-20675
Sequence 20675, Application US/09252991A
Parent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20675
LENGTH: 726
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20675

Alignment Scores:
Pred. No.: 5,91e-10 Length: 726
Score: 237.00 Matches: 210
Percent Similarity: 32.21% Conservative: 57
Best Local Similarity: 25.33% Mismatches: 279
Query Match: 5.61% Indels: 283
Gaps: 47

US-10-028-952a-4 (1-2333) x US-09-252-991A-20675 (1-726)
QY 2161 CCCCACTTCCATAGCCAGCAGAGCAGAGAAATGAGTGGGAGCAGAGCAGG 2102
Db 3 ProSerPhePro-----ArgArgSerArgAlaLeuGlu 13
QY 2101 CTCACACAGACAAATTCCTGCTGCCCAACACCAATGATCACTGACTTGTGTACA 2042
Db 14 ---ProLeuAspGluHisGlyLysThrInPro----- 23
QY 2041 AACTGTGTAATAAACAATTCCTAGCTTCACCTGTTCCAGAGGTCAATTCMAACAGTGT 1982
Db 24 -----SerAsnArgArgLeuArgArgLeuGluAlaArgAla-----AspAla 37

```

```

QY 1981 GGCCACCAACCCCTTAGAA-----GCAAGAGAAAAAGCAAAATAATGCAGAA 1928
Db 38 GlyAsnHisProLeuSerArgLeuAlaGlyAlaGlnProAlaGluGlnProGlyArg 57
QY 1927 GGCACGCCAGAGCCCTCCAGATGCTGCTGACAGACCTTCCCGCAGCTGGAGAGTGGC 1868
Db 58 GlyGlnAlaArgThrArgProAlaProThr-----ThrArgGlyProHis 72
QY 1867 CACTCAGAGCTTGGCCACAGAGGAGGCTTGCTCATGATCTCTGACGGGCGCAAGACCA 1808
Db 73 HisProGlyLeu---ArgArgArgValLeuAla----- 82
QY 1807 GCTGTCTGGAGAGGTCGGGGAGTGCAGAGAGATGTGTGAGATACCAAGGCCATTGCTG 1748
Db 83 -----TrpGlnAspArgAlaAspGlnGlnProValLeuArgAlaGlyLeuArgPro----- 98
QY 1747 CATTGGGGGTGACAGGCTGAGTGTGCGAAGCCGAGGAGGCTGAGGCTCATCTCAC 1688
Db 99 -----AlaHisAlaAlaValProArgArgAlaHisSerHisValProHis 113
QY 1687 TGGGGG---TCCTATTCCTGGAGAGTGTGATGTCAATCCCGAGGCTGCTCTTCAG 1631
Db 114 ArgAlaValProArgSerAlaArgArgAlaGlnLeuHisProPro-Ala-----Al 130
QY 1630 CAATAGAGAACCCATGTGTGAGAGGGAGG-----GCAAGATGTGCC----- 1588
Db 130 AspArgSerAlaProGlyArgArgGlnArgProValGlnAlaArgAlaAlaProLe 150
QY 1587 -----CTCAGCCCTCAGTCATGAGCTCCA----- 1564
Db 150 uGlyGluHisSerAlaGlyHisSerArgProGlyGlyHisGlyProGlyLeuArgProGly 170
QY 1563 -CGTAGTGGCAGGAGACATGC---CAAAATGCCATCCGGCCCATACGCCACGACGAG 1508
Db 170 YArgGlnAspGlnGlyAlaAlaGlyGlyThrGlyAspProAla----- 184
QY 1507 CCTTGTGATCACTTCAGTCCGCTGATGAGAGTCTCGGGGTCAAGAGATCTCTGTG 1448
Db 185 ---ArgLeuProProGlyCysPro-----GlySerAlaArg----- 195
QY 1447 TGTGCGGCTGCTGTAGTGTGACAGG-----CACGGGACAGAGCCCTTGC 1400
Db 196 -ArgAlaArgProGlyAlaArgProGlyLeuAlaProArgProGlyGlnLeuArgProArg 215
QY 1399 CCAGTAGCCCTTGGC-----CCTGAATGTGTGTGATG 1364
Db 215 AlaAlaAlaProGlyProAlaAspProArgHisAlaTrpProGluCysProArgGlnArg 235
QY 1363 TGTCCAGAGCCAGACCTTGTGCTGCA-----CCAGTGGGGGCTGCTCG 1319
Db 235 GAlaGlyAlaAspProValAspAlaAlaGlnArgProGlyAspProValProAlaValArg 255
QY 1318 TAGAAG-----TCTCTGCTCTGGA 1298
Db 255 GArgArgArgCysHisArgLeuGlyHisGlyTyrLeuAlaGluValHisProAlaThrAl 275
QY 1297 GGTTCCTATACAGACCTTCTGCTGCTGACCAAGC----- 1258
Db 275 ArgGlyArgProGlnGlnProValArgArgAlaGluGlnAspArgArgAlaLeuGlyArg 295
QY 1257 -----ATGAGAGTGTGCGCGCGGCTCTTCAGCAGGAGATCT 1217
Db 295 gProPheArgArgAlaLeuArgThrGluArgHisProProAspProGlyArgHisGlyArg 315
QY 1216 GCCCTGGGCTTGAGATGGCAGAGCTGCTGCTGCCAAAGTGGAGCTGCTGCTGGGTG 1157
Db 315 oPro-----AlaArgProAlaAlaArgArgArgAlaAlaValGlyGly 330
QY 1156 ACTGCTTTCAGAGAGAGGCTCTCAGCTTGCAGGCTGACGACTGAGATGGAGGTG 1097
Db 330 uAlaGlyAlaAlaGlyGlnGly-ProArg-----ArgThrGlyAlaAlaArgT 346

```

[illegible]

```

Db      623 rolaAGlnPFGDGLYLeuArgArgValArgArgGlyIleAlaAGlyThrValAlaArgSerA 643
Qy      89 -----GGGTTCTTTCCCGCGGATTCGCCCA 62
Db      643 rGArgHISProProProValAlaProProLeuAlaProAlaAlaProGlyGlnGlyG 663
Qy      61 AGCCCGTTCCTTGAGT 45
Db      663 InProAlaArgLeuAla 668

RESULT 11
US-09-252-991A-32367
; Sequence 32367, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32367
; LENGTH: 1418
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32367

Alignment Scores:
Pred. No.: 7.78e-10 Length: 1418
Score: 236.50 Matches: 195
Percent Similarity: 33.02% Conservative: 53
Best Local Similarity: 25.97% Mismatches: 249
Query Match: 5.59% Indels: 255
DB: Gaps: 40

```

```

US-10-028-552A-4 (1-2333) x US-09-252-991A-32367 (1-1418)
QY      1931 AGAGGGACGCAGGCGCTCCCAAGATGCCCTGTCCAGGAGCOCCTTCCCC----- 1884
        |||||
Db      10 ArgArgInProAlaPyrProLeuArgInGlnHisAlaIalAlaIalProSerAlaAsp 29
        ArgArgInProAlaPyrProLeuArgInGlnHisAlaIalAlaIalProSerAlaAsp
QY      1883 -----GCAGCTTGCGCATGTGGCCACTCAGGGCTTTGGCCACAGGACGGCGTTGGCTCA 1833
        |||||
Db      30 AlaSpheAlaAlaValGlnLeuProValArgAlaGlnProArgGlyPro----- 46
        |||||
QY      1892 GTATTCGTGACGGGGCAGAAGCCAAAGCTGTGTGGAGGATCGGGGATGCAGAAGATG 1773
        |||||
Db      46 ----- 46
QY      1772 TGTGGGATCACAGGCCATTGTGTGATTGGGGTGCACAGGCTGAHTGTCCAAAGCCGGA 1713
        |||||
Db      47 -----ArgAlaPro 49
        |||||
QY      1712 GGGAGCCCTGAGGCGCTCATCTCTACGTGGGGTCTTATTCTCGAAGAGAGCTGTAATGCA 1653
        |||||
Db      50 GlyAlaProAlaGmTrAlaArgSerLeuGly----- 59
        |||||
QY      1652 ACTCCCAGGCGCTCTCTTCCAGCATTAAGGAAGCCATGTCTGAAGAGGGGAAGGCCAAGAT 1593
        |||||
Db      60 -----ArgProGlnAlaIalLeuArgInProGlnHisArgArgLeuGly----- 74
        |||||
QY      1592 GTGCCCTCAGCGCTACTCATATGAGCTCCAGCTAATTGTGGAGGGAACATGCCCAAATGACC 1533
        |||||
Db      75 -----HisAlaAspHisArg-HisLeuProAlaTrAl 85
        |||||
QY      1532 ATCCGGCCCATAGCCAGC----- 1514
        |||||

```


Db 85 aalaglyargheuproargalaalaargalaglhlhlararglalnargheupheglly 105
 QY 1513 ----CACAGGCTTCGATGATCAGATGAGCCGATGAGGTTCTGGGGTCAA---- 1462
 Db 105 ygllyarglhnargllyalalaspargargproalaleuargleu-Argllyglhly 125
 QY 1461 ----AGAGATCTCTGTGT--CGTCGGCTGCTGTGTAGTCGTACAGGGACGGGACCA 1410
 Db 125 erValargargalaleuengllyarglaalaprogllyserargarg-----HlsA 142
 QY 1409 GAGCCCTTGGCCCATGAGCCCTGGCCGTGAATGTGTGTCATGTGTCTCAGAGCCAGC 1350
 Db 142 rgleuargalavalaglhnleuargproprocyS-----Alalagllyalargh 159
 QY 1349 ACCTGTGCTGTCACACAGCTGGGGCTGCTCGTAGAG-----TCTCTGCTC 1302
 Db 159 isProargllepProarganProargProleuylasProserH:gluProserL 179
 QY 1301 TGGAGGTCCTCATACAGACCTCTCTTCTGCTGACCCAGACATGAGAGAGCTGGG 1242
 Db 179 euargalargsertharglnProhlshispro-----Ttp- 191
 QY 1241 CGCCGCTCTCTCAGAGGAGATCTGCTGCTGGCC-----TTGAGATGGCAGCAGC 1191
 Db 192 -----GlnargProasleuProgllyalalagllyleuargargproglhl 208
 QY 1190 TGGCTCTTCCCAAGTGGTCTCTGCTG-----GGTGAAGCTCTCT 1149
 Db 208 euargargProargProgllyasn-glyleuargserargalalavalaleu 227
 QY 1148 CTGCAG-----GAAGGGCTCTCAGCTTCCAGCTGACGCTGAGACTGAGATGAGT 1098
 Db 228 leuaglalnarglyargargglyPro-----glyargargglhnlhserglYalA 245
 QY 1097 GATTAGGCGCAGGTTCCCAAGACGTGCGCTGTCAGCGGAGAGGGCGCGCCCT 1038
 Db 246 glyleuPro-gluProaspararglnleuProleuargspargThrglygllyaspar 265
 QY 1037 TTCCGGCGCGCCGCTTCC-----AGAGCAAGAGG 1005
 Db 265 gglnglyargleuarghlhserProgllyalaglYarglythargleuargleuargl 285
 QY 1004 CACTCCG--CACCGAGCCCGGCTCCCGCGCGCGCGCGCGCGCCCTCCCGCGCGC 948
 Db 285 yglYProgllyhlsapProargProalaspalahlhlsalargspglPro--Al 304
 QY 947 GCGGGCGCGTGGAGGGGGGGCGCGCGCGCGCGCGAGCAGCGGGGACCGGC--TA 891
 Db 304 aleuaglasparglnnglyleuargProleuaglYalalaspargProalaleuProcySgl 324
 QY 890 TCCGAGGCCAACCGAGGCTCCCGCGCGCTCGTATCGTTCCGCTGGCGGATTTCTGA 831
 Db 324 nProgllyargProseralaproargarghlhln-----ProProgllyargargargl 342
 QY 830 CTTAAGGCGCTCATCATATCCACAGATGTGTGCTCCGCCCATGCTCTCTCAGCC 771
 Db 342 yserProalaglalnarghlhargProvalasp-----leu-ValgllyglYasn 358
 QY 770 AAGCATATACACCAATGTCTGAACCTGCGTCTCTCTGATGAGCAGATTACC-- 714
 Db 359 Arglnleu-----Provalalaleuargalalapro 370
 QY 713 ---ATGGCAACAACATCATCATAGTAGGTAAATACTACCTGTCTCAGACGAGTATC 657
 Db 371 ArgargProgllyleuhlharghlhserProgllyalalaseralalargseralalaser 390
 QY 656 CCAGCTCAGTCCCTATTAGTGGTGAACAATCCACGCTGTGTAATCTGCTTCA 597
 Db 391 GlyYseralapro----- 395
 QY 596 ATGATAGGAAGCGCAGCATCGAAGATCAAAAGCAGCGTGTGATGACGCTTGG-- 540
 Db 396 -----Serthralaseralalaseralaseraltper 407

QY 539 -----CCGCGCAAA-----GCCAGTTATCCCTGTGTAACTTTCTGACACCTCTGC 492
 Db 408 AspProProalahlh:glYalaglYProProcyS-----ProserSerly 423
 QY 491 TTTAAACCCMAAGTCAAGAGATCGTGAGGCCCCGCTTACGCTGTGATTTGACT 432
 Db 424 leuSerProargargargProargasn--AlahlalargYalA-----Argln 439
 QY 431 GAAATCAAGATCA-----GGAGCTTTGCCCTTCTGCT 396
 Db 440 GlnglnglnaspargargProgllyphenlsargysalaglYglYphenargleuProval 459
 QY 395 CCAGGAGGTTTCTGTCTCTCTCTAGCTGCGCTTACACACCTGCTTAC--CGTTG 339
 Db 460 AlaserProvalalaleuProaspararghlhlsiglYargllyleuarghlharghlh 479
 QY 338 ACAGGTGACCGCCCAAGTCAAACTCCACCTGCGACTGTCCCGGAGCGGGTGCAGC 279
 Db 480 -ArgleuhlhargProgllyProgllyProargleuahlhngluProgllyalathrglnPr 499
 QY 278 CGGC--CGCGCGCGCGCGCGCGCTTGGCGCCAGAA-----GGAGAGCCCTCGGG 228
 Db 499 asPProargargargalaglYargProgllyProargargleuargargargProalagl 519
 QY 227 CTCGCCCCCGCC-----CTCACCGGCTCAGTGAATAAAGATCAGA 186
 Db 519 yargProphenargthnglngluargalaproalaleuylthValleuargargleu 539
 QY 185 GTA-----GTGATTTTCACCG 169
 Db 539 UPProargleuaglYleuargargphenlyglYasnProgllyProalalaleuPhnlharg 559
 QY 168 GCGGCGCGCGCGCGCGCGAG-----CCGCGC 142
 Db 559 gProargProgllyargglYasparaglalnargspHlsProasleuarglileuProal 579
 QY 141 CCGGGCGCC--TGCGGGGACACCGGGCG-----GGCGC 109
 Db 579 aProalaproargalalaspglYhlhargspvalleuargleuHlsProargleuaglYal 599
 QY 108 CGGG-----GGCCAACTCAACAGGCTTCTTCCCGCTGA 70
 Db 599 aglyargProalalargargalalaspolyProgllyalaglYleuaglYlsargalalagl 619
 QY 69 TTCGCGCAACCCGTTCC 51
 Db 619 aArgargargProalapro 625

RESULT 12
 US-08-942-423-4
 / Sequence 4, Application US/08942423
 / Patent No. 5891673
 / GENERAL INFORMATION:
 / APPLICANT: Hashimoto, Yasuhiro
 / APPLICANT: Takemoto, Yoshihiro
 / TITLE OF INVENTION: Lock Binding Protein
 / NUMBER OF SEQUENCES: 68
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Synex (U.S.A.) Inc.
 / STREET: 3401 Halliwell Ave.
 / CITY: Palo Alto
 / STATE: California
 / COUNTRY: U.S.A.
 / ZIP: 94303
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patentin Release #1.0, Version #1.30
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/942,423
 / FILING DATE: 01-OCT-1997

Fri Apr 2 07:35:03 2004

us-10-028-952a-4.rail

Page 13

[illegible][illegible]

```

Db 185 sGIyArgProProAlaCysSerProValAlaThrAlaProThrAlaThnCysSerProR 205
QY 1441 -----GTCGCTGTATGTCGTACAGGACGAGGACGAGCCCTTGC----- 1400
Db 205 oSerAlaArgSerAlaTrp-----LysProCysAlaCysAl 217
QY 1399 -----CCACTGAGCCCTGCGCC-----TGAATGTGTG 1371
Db 217 alyAlaValAlaGlySerAlaArgSerProAlaSerAlaTrpProAlaAlaSerAlaTrpC 237
QY 1370 GTCAATGTGCTACAGGACGAGCCTTGTCTGTCACAGGAGGAGGCTGCTGTAGAGAGT 1311
Db 237 sProThrCysAsnSerAlaArgPro-SerAla-----IleAlaSerThrAlaAlaArgS 255
QY 1310 CTCTGCTCTGGAG-----GTTCTCATACACAGCCTCTCTTCTGCTG 1266
Db 255 expSerThrAsnGlyArgProAlaAlaCysProThrThrThrProPro----- 271
QY 1265 CACGACAGATGAGAGAGTGTGGCGCGGCTCTCAGCAGGAGAGATCTGCGGAGCT 1206
Db 272 -----AlaAsnAlaProAlaSerAlaAlaAlaAlaAlaAlaAlaAlaAla 284
QY 1205 TGAGATGGCAG-----CAGCTGCTCTCTGCCAAGTGGTCTGTGTGGGTGAGCTG 1152
Db 285 -----TrpLysAlaProSerProAlaMetProProSerThrSerAlaThrArg---LeuA 302
QY 1151 CTCTGTGAGAGAGGAGGCTCTCCTCAGGCTGCGAGCTGAGACTGAGATGAGAGTGTGTTA 1092
Db 302 IaAlaCysGlyHnAspGlyAspAlaGlyAlaSerGlnThrProTrpArgAspThrS 322
QY 1091 GCGGCGAGTTCCCGACGACATG-----CGTGCGTGACGCGGCGAGGCGGCGCGG 1041
Db 322 expProCysValProAsnGlyValAsnTrpArgArgGlyLeProTrpSerAlaValAla 342
QY 1040 CCTTCCGCGCGCGCGCTTCCAGAGACGAGAGGACCTCCGACCGCGCGCTG-- 983
Db 342 ystHiserProTrpProArgTrpProGlyMetProAla-----GlyProArgProS 359
QY 982 -----CCGCGCGCGCGCGGCGGCGACGCGCGCTCCGCGCG- 950
Db 359 erCysThrAlaValAlaAlaAlaTrpAlaProAlaArgGlyGlyTrpAlaAlaAlaProArg 379
QY 949 -----GCGCGCGCGCGCTGGA 934
Db 379 snSerCysSerAlaGlyArgLeuArgArgValAlaArgArgSerAlaArgArgAlaTrpA 399
QY 933 GG-----GGGCGCGCG-----CCGCGCGCGGAGACAGGCGGAGCCGCTATC 889
Db 399 rgProTrpProAlaGlyArgAlaThrProAlaSerAlaArgArgAspSer-AlaSer 418
QY 888 CGAGGCGCACCGAGGCTCCGCGCGCTGCCGTATCTCCGCTGGCGGAGATTCTGACT 829
Db 419 ArgGlyArgPro----- 422
QY 828 TAGAGCGGTCAGTCATATCCACAGATGTAGCTTCCGCCCAT-----TGAGTC 778
Db 423 -----AlaThrAlaAlaAspHisProAlaAlaAlaTrpVal 433
QY 777 CTCGACCAAGACATACCAAAATGTCTGAACCTGCGGTTCC-----TTCGTACT 727
Db 434 SerAlaAlaArgArgThrSerSerAlaProIleAlaGlySerAlaProGlySerGlyThr 453
QY 726 GAGCAGAGATTACATGCAACACACATCATCAGTAGGTAAACTAACCTGTCTCAGCA 667
Db 454 AlaProArgCysHis-----ProValArgGlys 462
QY 666 ---CGTCTAATCCAGCTCAGTTCCCTATTAAGTGGGTGAACAATCAAGCTTGCTGA 610
Db 463 AspGlyAlaGlyProAlaSer-----ThrGlyArgSerArgArgArgTrp--- 477
QY 609 ATTCTGCTTCAATGATGATAGAGAGCGGA----- 580
Db 478 ---AlaAlaAlaProAlaArgArgAlaArgThrGlyIleArgArgThrAlaAlaAla 496

```

```

QY 579 -----CATGAGAGATCAAAAAGCAGACGTGCTATGAACGC 544
Db 497 CysTrpProProAlaProGlyArgArgThrProAlaAlaAlaArgArgSerAlaArg 516
QY 543 TTGCGCGGACAGCGAGTTATCCCTGTGTAATTTTCTGACACCTCTGCTTAAAC 484
Db 517 ThrAla-ProArgProAlaAlaProAlaHis-----ArgPr 528
QY 483 CAAGAAGTCA--GAAGAGTGTAGAGGCGCGCTTCAACGCTGTATTGTAAGAAA 427
Db 528 cSerAlaSerAlaThrGlyTrpProGlyProAlaLeuAla----- 541
QY 426 TCAAGATCAAGAGCTTTGCGCTGCTCCACAGGAGGTTCTGTCCTGCTGAGCT 367
Db 542 -----CysProAlaAlaAla-GlyArgArgProSer-Pro---Ala 553
QY 366 CGCCTTAGACACCTGCGTTACCGTTGACAGGTACCGCGCCAGTCAAACTCCAC 307
Db 554 ProAlaProAlaProAla--ProHisAlaArgProAlaArgAlaProArgArgPro--- 571
QY 306 TGGCACTGCTCCCGAGCGGCTGCGCGCGCGCGCGCGCGCGCGGCGCTTGCGGCG 247
Db 572 -----GlyArgSerProAlaThrAla--ProAlaAlaAlaAla 584
QY 246 AAGCAGAGACCCCTGCGGCTGCGCGCGCGCGCGCTCACCGGAGTCAAGAAAACATCAG 187
Db 585 ProAlaGlyProThrProAlaArgProProAlaAlaAla-----ThrGlyArgPro 601
QY 186 AGTACTGTATTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 145
Db 602 SerAlaProProArgProAlaArgArgArgValArgTrpProTrpArgSerAlaArgCys 621
QY 144 -----CGCGCGCGCGCGCTGCGCGGAGCACCGCGGCGCGCGCGCGCGCGCGAGCTC 94
Db 622 SerProArgArgProThrPro---ArgSerAlaPro-GlyAlaAlaThrGlyAlaProTh 640
QY 93 AACAGGCTCTTCTTCCCGCTGATTCCGCGCAACCGCTTCCC 51
Db 640 rThrGlyAlaAlaArgProArgArgArgAlaAlaAlaPro 654

```

Search completed: April 1, 2004, 15:59:31
 Job time : 53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 1, 2004, 15:47:41 ; Search time 117 seconds
(without alignments)
12582.974 Million cell updates/sec

Title: US-10-028-952a-4

Perfect score: 4230

Sequence: 1 gcaagagatccaccatgctcc.....agcttggaactctcttcgcca 2233

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
	Xgapop 10.0 , Ygapext 0.5
	Fgapop 6.0 , Fgapext 7.0
	Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool.p/US10028952/runat.01042004.154625.652/app.query.fasta_1.2503
-DB=SPTREMBL_25 -QWMT=faстан -SIFFIX=resp -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdd -LIST=45
-DOCALLIGN=200 -THR.SCORE=spct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028952@cgn 1.1.128 @runat.01042004.154625.652 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_25:*
1: sp.atchea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mmc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvins:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	877	20.7	430	4	Q9NR72	Q9nr72 homo sapien

2	872	20.6	302	4	Q9HBN8	Q9hbn8 homo sapien
3	872	20.6	430	4	Q9JUT6	Q9jut6 homo sapien
4	870	20.6	431	4	Q9CF30	Q9cf30 homo sapien
5	870	20.6	433	4	Q96K74	Q96k74 homo sapien
6	759.5	18.0	282	10	Q9AVH2	Q9avh2 pium sativ
7	703.5	16.6	436	11	Q8OWP1	Q8owp1 mus musculu
8	702.5	16.6	433	11	Q62418	Q62418 mus musculu
9	694.5	16.4	432	11	Q8BH56	Q8bh56 mus musculu
10	665.5	15.7	436	11	Q9JH14	Q9jh14 rattus norv
11	665	15.7	429	11	Q9JH74	Q9jh74 rattus norv
12	661.5	15.6	432	11	Q9JH74	Q9jh74 rattus norv
13	661	15.6	433	11	Q9JH74	Q9jh74 rattus norv
14	392.5	9.3	447	13	Q7ZXQ9	Q7zxq9 xenopus lae
15	288.5	6.8	643	5	Q9XUT0	Q9xut0 caenorhabd
16	265.5	6.3	384	5	Q8MR59	Q8mr59 dirosophila
17	265.5	6.3	531	5	Q9YU84	Q9yu84 dirosophila
18	261.5	6.2	486	11	Q922T8	Q922t8 mus musculu
19	261	6.0	559	5	Q9YDP4	Q9ydp4 dirosophila
20	255	6.0	587	5	Q96459	Q96459 strongiloe
21	253.5	6.0	559	5	Q96046	Q96046 dirosophila
22	243	5.7	270	10	Q64410	Q64410 zea mays (m
23	236	5.6	471	11	Q70419	Q70419 rattus norv
24	236	5.6	508	11	Q70420	Q70420 rattus norv
25	233	5.5	477	5	Q96620	Q96620 suberites d
26	233	5.5	513	4	Q96H99	Q96h99 homo sapien
27	231	5.5	509	11	Q921L6	Q921l6 mus musculu
28	231	5.5	509	11	Q8BNAS	Q8bnas mus musculu
29	231	5.5	530	13	Q8UWC3	Q8uwc3 xenopus lae
30	220.5	5.2	774	12	Q41971	Q41971 murid herpe
31	213.5	5.0	706	12	Q41972	Q41972 murid herpe
32	213.5	5.0	727	12	Q41973	Q41973 murid herpe
33	209	4.9	890	12	Q8UZE1	Q8uze1 cercopithec
34	206	4.9	206	13	Q7SXH4	Q7sxh4 brachydanto
35	206	4.9	1478	5	Q961U4	Q961u4 dirosophila
36	206	4.9	1843	5	Q9V6W9	Q9v6w9 dirosophila
37	204	4.8	1463	16	Q9ADP6	Q9adp6 streptomyce
38	203.5	4.8	608	12	Q9Q5K9	Q9q5k9 herpesvirus
39	201.5	4.8	1111	16	Q9L0P1	Q9l0p1 streptomyce
40	200.5	4.7	758	5	Q86KTE	Q86kte dictyostell
41	199	4.7	530	10	Q04892	Q04892 nictotiana t
42	198.5	4.7	634	4	Q8N707	Q8n707 homo sapien
43	198	4.7	745	16	Q89X06	Q89x06 bradyrhizob
44	196	4.6	579	2	Q86FF2	Q86ff2 streptomyce
45	194.5	4.6	368	3	Q9P837	Q9p837 candida alb

ALIGNMENTS

RESULT 1
Q9NR72 PRELIMINARY; PRT; 430 AA.
ID Q9NR72
AC Q9NR72
DT 01-OCT-2000 (TREMURel. 15, Created)
DT 01-OCT-2000 (TREMURel. 15, Last sequence update)
DT 01-OCT-2003 (TREMURel. 25, Last annotation update)
DE Cervical SH3P7 (Mucin-associated protein).
GN CMAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ho S.B., Toribara N.W., Anway R.E., Spurr-Michaud S.J., Shekels L.L.,
RA Keutmann H.T., Hill J.A., Gibson I.K.,
RT "Expression cloning of human cervical proteins using an antibody to
RT cervical mucus";
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RX MEDLINE=99077625; PubMed=9858486;
RA Gibson I.K., Spurr-Michaud S., Moccia R., Zhan Q., Toribara N.,

RA Ho S.B., Garzino A.R., Hill J.A. III,
RT "MCA and MCB transcriptions are the prevalent mucin messenger
RL ribonucleic acids of the human endocervix.";
CC Biol. Reprod. 60:58-64(1999).
DR -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF250287; AA81273.1; -.
DR EMBL; AF151364; AA3120.1; -.
DR HSSP; P23727; 1PNU.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; Cofilin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 430 AA; 48154 MW; 7D0CA17D8CB8DC CRC64;

Alignment Scores:

Pred. No.:	1,32e-56	Length:	430
Score:	877.00	Matches:	180
Percent Similarity:	74.60%	Conservative:	8
Best Local Similarity:	71.43%	Mismatches:	56
Query Match:	20.73%	Indels:	9
	4	Gaps:	2

US-10-028-952A-4 (1-2333) x Q9NR72 (1-430)

QY 841 GCCCAGCGGAGACGATACGAGCGCGCGAGCCTCGTGGCTCGATACCGCTCC 900
DB 181 AlAlYsAlAGlulYs-----GlulGluluaInArgrgluGlulYsArgrAla 198
QY 901 CCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 960
DB 199 GlulGlulAGlulArgrgluGlulGlulGlulArgrgluArgrgluAla 218
QY 961 CGAA 1020
DB 219 ArgrArgrgluInArgrgluInglulGlulYsArgrgluInglulYs 238
QY 1021 ACGGG-----GCCGCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCTCA 1061
DB 239 GlulGlulGlulValValSerArgrArgrgluInglulGlulYsArgr 258
QY 1062 CGCAGCGACGCTTCTGGGAGACCTGGCGCTAAACACCTCCATCTCCAGTCCAGCT 1121
DB 259 GlulPhePheGlnYsGlnYsGlnYsGlnYsGlnYsGlnYsGlnYsGlnYs 278
QY 1122 GSCAAGCTGAGAGCGCGCTTCTCGAGAGAGCTCACCAACAGAGAGCGCTTGGC 1181
DB 279 GlYlYsLeuArgrSerProPheLeuGlnYsGlnYsGlnYsGlnYsGlnYs 298
QY 1182 AAGAGCGAGCTGCTGCTCAAGAGCGCGAGATCTCCCTCGTGAAGAGCGCG 1241
DB 299 ArgrgluProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 318
QY 1242 CCAGAGCTCTCCATGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1301
DB 319 ProSerThrProProCysLeuValGlnAlaGlnYsGlnYsGlnYsGlnYs 338
QY 1302 GAGCAGAGAGAGCTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1361
DB 339 GlulGlulGlulThrPheYrGlulInProPheLeuValGlnYsGlnYsGlnYs 358
QY 1362 CACATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1421
DB 359 HisIleAspHisIleIleGlnYsGlnYsGlnYsGlnYsGlnYsGlnYs 378
QY 1422 CTGTACGACTACGAGCGAGCGAGCAGCAGAGATCTCTTGAACCGAGAGAGCTCATC 1481

DB 379 LeuYrAspYrGlnAlaAlaAspThrGlnLeuSerPheArProGlnYsLeuIle 398
QY 1482 ACGGCGATGAGATGATGAGAGAGCGTGGCGTGGCTATGAGCGAGAGCGCATTTT 1541
DB 399 ThrGlyIleGlnValIleAspGlnYrPTrpArgrgluYrGlyProAspGlyHisPhe 418
QY 1542 GGCATGTCCTCGCAACTAGTGAAGCTCATTTAG 1577
DB 419 GlyMetPheProAlaAspYrValGlnLeuIleGln 430

RESULT 2
Q9HNB8
ID Q9HNB8 PRELIMINARY; PRT; 302 AA.

AC Q9HNB8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.X., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human CDNA clones with function of inhibiting cancer cell
RT growth." (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF218020; AAG17262.1; -.
DR HSSP; P23727; 1PNU.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Hypothetical protein; SH3 domain.
SQ SEQUENCE 302 AA; 34388 MW; 2DECB1A4C7511612 CRC64;

Alignment Scores:

Pred. No.:	2.97e-56	Length:	302
Score:	872.00 <td>Matches:</td> <td>179 </td>	Matches:	179
Percent Similarity:	74.21% <td>Conservative:</td> <td>8 </td>	Conservative:	8
Best Local Similarity:	71.03% <td>Mismatches:</td> <td>57 </td>	Mismatches:	57
Query Match:	20.61% <td>Indels:</td> <td>9 </td>	Indels:	9
	4 <td>Gaps:</td> <td>2 </td>	Gaps:	2

US-10-028-952A-4 (1-2333) x Q9HNB8 (1-302)

QY 841 GCCCAGCGGAGACGATACGAGCGCGCGAGCCTCGTGGCTCGATACCGCTCC 900
DB 53 AlAlYsAlAGlulYs-----GlulGluluaInArgrgluGlulYsArgrAla 70
QY 901 CCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 960
DB 71 GlulGlulAGlulArgrgluGlulGlulGlulArgrgluArgrgluAla 50
QY 961 CGCGTCCCGAA 1020
DB 91 ArgrgluInArgrgluInglulGlulYsGlnYsGlnYsGlnYsGlnYsGlnYs 110
QY 1021 ACGGG-----GCCGCGCGGAGGAGCGCGCGCGCGCGCGCGCGCTGCA 1061
DB 111 GlulGlulGlulValValSerArgrArgrgluInglulGlulYsArgr 130
QY 1062 CGCAGCGAGCTTCTGGGAGACCTGAGCTTAACACCTCCATCTCCAGTCCAGCT 1121
DB 131 GlulPhePheGlnYsGlnYsGlnYsGlnYsGlnYsGlnYsGlnYsGlnYs 150

QY 1122 GGCAAGCTGAGAGACCCCTTCTGACAGAACAGCTCCACCCAGACCACTTGGC 1181
 DB 151 GILYLSLEUATGSEPRPHELEUGLNLGLNLEUTHRGLNPROGLNTHRHISPHGLY 170
 QY 1182 AGAGAGCCAGCTGTGCTCATCTCAAGGCCCAAGGCGACATCTCCCTGTGAGAGCCGGG 1241
 DB 171 ArgGluProAlaAlaAlaAlaIleSerArgProArgAlaAspLeuProAlaGluGluProAla 190
 QY 1242 CCAGCACTCCTCCATGCTGTGTCAGCAGAGAGAGGCTGTATGAGAACCTCCA 1301
 DB 191 ProSerThrProProCysLeuValGlnAlaGluGluGluAlaValTyrGluGluProPro 210
 QY 1302 GAGCAGAGACCTTCTCAAGACAGCCCTGCTGAGCAGACAGAGTGTGCTGTAG 1361
 DB 211 GlnGlnGlnThrPheTyrGlnGlnProProLeuValGlnGlnGlnGlnGlnGlnGln 230
 QY 1362 CACATGACCAACCATTCATGAGGCGAGGCTCAGTGGCAGGCTCTGTGCCCCGTGCC 1421
 DB 231 HistLeaspHisHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 250
 QY 1422 CTGTACACTACACGAGCAGCAGCAGCAGACAGAGATCTCTTTGACCCCGAAGCTTCATC 1481
 DB 251 LeuTyrAspTyrGlnAlaAlaAlaAspAspThrGlnIleSerPheAspProGluAsnLeuIle 270
 QY 1482 ACGGGCATCGAGGTGATCGACGAGAGGCTGGTGGCTATGGGCGGATGGCATTTT 1541
 DB 271 ThrGlyIleGluValIleAspGlnGlyIlePheArgGlyTyrGlyProAspGlyHisPhe 290
 QY 1542 GGCAATGTTCCCTGCAACTACGTGAGCTCATTTAG 1577
 DB 291 GlyMetPheProAlaAsnTyrValGluLeuIleGln 302
 RESULT 3
 Q9JUT6 PRELIMINARY; PRT; 430 AA.
 ID Q9JUT6
 AC Q9JUT6
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DE SRC homology 3 domain-containing protein HTP-55 (Drebrin F).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20035526; PubMed=10567356;
 RA Emsenat D., Yao Z., Wang X.-S., Kori R., Zhou G., Lee S.C., Tan T.-H.;
 RT "A Novel Src Homology 3 Domain-containing Adaptor Protein, HTP-55,
 RT that Interacts with Hematopoietic Progenitor Kinase 1.";
 RL J. Biol. Chem. 274:33945-33950(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Zhang W., Yuan Z., Man T., He L., Cao X.;
 RT "Molecular cloning of cDNA encoding drebrin F";
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Straubeberg R.;
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 DB EMBL; AF137060; AA13701.1;
 DB EMBL; AF077353; AA80228.1;
 DB EMBL; BC031687; AA831687.1;
 DB HSSP; P23727; 1PNU.
 DR GO; GO:0008047; F:enzyme activator activity; TAS.
 DR GO; GO:0005515; F:protein binding; TAS.
 DR GO; GO:0002257; P:activation of JNK; TAS.
 DR InterPro; IPR002108; Acbbind_cofin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; coflin_ADP.1.
 DR Pfam; PF00018; SH3.1.
 DR PRINTS; PR00452; SH3DOMAIN.

DR Prodom; PD000066; SH3; 1.
 DR PROSITE; PS50002; SH3; 1.
 SQ SEQUENCE 430 AA; 48207 MW; 7E8C42ED047257AE CRC64;
 Alignment Scores:
 Pred. No.: 3 11e-56 Length: 430
 Score: 872.00 Matches: 179
 Percent Similarity: 74.21% Conservative: 8
 Best Local Similarity: 71.03% Mismatches: 57
 Query Match: 20.61% Indels: 9
 DB: 4 Gaps: 2
 US-10-028-952a-4 (1-2333) x Q9JUT6 (1-430)
 QY 841 GCCACGCGGAAACGATACGCGACCGCCGCGAGCTCGTTGGCTCGGATAGCCGTTCC 900
 DB 181 AlAtysAlaGluTyr-----GluGluGluAsnArgArgLeuGluGluTyrArgArgAla 198
 QY 901 CCGGCTGTCCCGCGCGGCGGCGCGCCCGCCCTCCACGCGCGCGCGCGGAGAG 960
 DB 199 GluGluAlaGlnArgGlnLeuGlnGlnGlnArgArgGlnArgGlnLeuArgGluAla 218
 QY 961 CCGCTGCCCG 1020
 DB 219 ArgArgGluGlnArgTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 238
 QY 1021 ACGGG-----GCGGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1061
 DB 233 GlnGlnGlnGlnValValSerArgAsnArgAsnGlnGlnGlnSerAlaValHisProArg 258
 QY 1062 CCGACCGGACGTTCCGTGGGGAACCTGGCGCTAAACACCTCATCTCCAGTCTCAGCTT 1121
 DB 259 GluIlePheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 278
 QY 1122 GGCAAGCTGAGAGACCCCTTCTGACAGAACAGCTCCACCCAGACCACTTGGC 1181
 DB 279 GILYLSLEUATGSEPRPHELEUGLNLGLNLEUTHRGLNPROGLNTHRHISPHGLY 298
 QY 1182 AGAGAGCCAGCTGTGCTCATCTCAAGGCCCAAGGCGACATCTCCCTGTGAGAGCCGGG 1241
 DB 299 ArgGluProAlaAlaAlaAlaIleSerArgProArgAlaAspLeuProAlaGluGluProAla 318
 QY 1242 CCAGCACTCCTCCATGCTGTGTCAGCAGAGAGAGGCTGTATGAGAACCTCCA 1301
 DB 319 ProSerThrProProCysLeuValGlnAlaGluGluGluAlaValTyrGlnGluProPro 338
 QY 1302 GAGCAGAGACCTTCTCAAGACAGCCCTGCTGAGCAGACAGAGTGTGCTGTAG 1361
 DB 339 GlnGlnGlnThrPheTyrGlnGlnProProLeuValGlnGlnGlnGlnGlnGlnGln 358
 QY 1362 CACATGACCAACCATTCATGAGGCGAGGCTCAGTGGCAGAGGCTCTGTGCCCCGTGCC 1421
 DB 359 HistLeaspHisHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 378
 QY 1422 CTGTACACTACACGAGCAGCAGCAGACAGAGATCTCTTTGACCCCGAAGCTTCATC 1481
 DB 379 LeuTyrAspTyrGlnAlaAlaAlaAspAspThrGlnIleSerPheAspProGluAsnLeuIle 398
 QY 1482 ACGGGCATCGAGGTGATCGACGAGAGGCTGGTGGCTATGGGCGGATGGCATTTT 1541
 DB 399 ThrGlyIleGlnValIleAspGlnGlyIlePheArgGlyTyrGlyProAspGlyHisPhe 418
 QY 1542 GGCAATGTTCCCTGCAACTACGTGAGCTCATTTAG 1577
 DB 419 GlyMetPheProAlaAsnTyrValGluLeuIleGln 430
 RESULT 4
 ID Q96F30 PRELIMINARY; PRT; 431 AA.
 AC Q96F30
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Similar to src homology 3 domain-containing protein HRP-55.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN NBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL: BC011677; AAL11677.1; -
 DR GO: GO:0005622; C:intracellular; IEA.
 DR GO: GO:0003779; F:actin binding; IEA.
 DR InterPro: IPR002108; Actbind_cofln.
 DR InterPro: IPR01452; SH3.
 DR Pfam: PF00241; cofilin_ADF; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00102; ADF; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KM SH3 domain.
 SQ SEQUENCE 431 AA; 48294 MW; D18FD316FFD0B4E CRC64;
 Alignment Scores:
 Pred. No.: 4,38e-56 Length: 431
 Score: 870.00 Matches: 161
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.57% Indels: 0
 Gaps: 0
 US-10-028-952a-4 (1-2333) x Q96F30 (1-431)
 QY 1095 ACCACCTCATCTCTCAGCTCCTCAGCTGCAAGCTGAGAGCCCTTCTTGCAAGAGCAG 1154
 DB 271 ThrThrSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnLysGln 290
 QY 1155 CTCACCCACACGAGAGACCCACTTTGGAGAGAGCCAGCTGCTGCACCTCAAGGCCACG 1214
 DB 291 LeuThrGlnProGlnThrHisPheGlyArgGlnProAlaAlaAlaIleSerArgProArg 310
 QY 1215 GCAGATCTCCCTGCTGAGAGAGCCGCCGCCAGCACTCTCCATGTCGTGAGAGAGAA 1274
 DB 311 AlaAspLeuProAlaGlnGlnProAlaProSerThrProProCysLeuValGlnAlaGln 330
 QY 1275 GAGAGGCTGTGTATGAGAGAACTTCCAGAGAGAGACCTTTCACGAGAGCCGCCACTG 1334
 DB 331 GlnGlnAlaValTyrGlnGlnProProGlnGlnGlnThrPheTyrGlnGlnProProLeu 350
 QY 1335 GTGAGAGCAGAGGTGCTGCTGAGACCACTTGACACCAACATTCAGGGCCAGGGGCTC 1394
 DB 351 ValGlnGlnGlnGlnAlaGlySerGlnHisIleAspHisIleGlnGlnGlnGlnGln 370
 QY 1395 AGTGGGCAAGGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1454
 DB 371 SerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 390
 QY 1455 ATTCCTTTCAGCCGAGACCTTCATCAGGGCATGAGGTGATGAGAGAGGTGAGGTG 1514
 DB 391 IleSerPheAspProGlnAsnLeuIleThrGlyIleGlnValIleAspGlnGlyTyrPtp 410
 QY 1515 CGTGGCTATGGGCGGATGCGCATTTGGCATGTTCCTGCAACTACGTTGAGAGCTTAT 1574
 DB 411 ArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGlnLeuIle 430
 QY 1575 GAG 1577
 DB 431 Gln 431
 RESULT 5

Q96K74
 ID Q96K74 PRELIMINARY; PRT; 439 AA.
 AC Q96K74;
 DT 01-DEC-2001 (TRENBLREL. 19, Created)
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
 DE Hypothetical protein FLJ1461.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN NBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Orsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
 RA Magatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Ninomiya K., Iwayanagi T.;
 RA "NEO human cDNA sequencing project";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL: AK027367; BAB55065.1; -
 DR GO: GO:0005622; C:intracellular; IEA.
 DR GO: GO:0003779; F:actin binding; IEA.
 DR InterPro: IPR002108; Actbind_cofln.
 DR InterPro: IPR01452; SH3.
 DR Pfam: PF00241; cofilin_ADF; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000066; SH3; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00102; ADF; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KM Hypothetical protein; SH3 domain.
 SQ SEQUENCE 439 AA; 49042 MW; BC044FBD0E82C6D5 CRC64;
 Alignment Scores:
 Pred. No.: 4,39e-56 Length: 439
 Score: 870.00 Matches: 161
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.57% Indels: 0
 Gaps: 0
 US-10-028-952a-4 (1-2333) x Q96K74 (1-439)
 QY 1095 ACCACCTCATCTCAGCTCCTCAGCTGCAAGCTGAGAGCCCTTCTTGCAAGAGCAG 1154
 DB 271 ThrThrSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnLysGln 290
 QY 1155 CTCACCCACACGAGAGACCCACTTTGGAGAGAGCCAGCTGCTGCACCTCAAGGCCACG 1214
 DB 291 LeuThrGlnProGlnThrHisPheGlyArgGlnProAlaAlaAlaIleSerArgProArg 310
 QY 1215 GCAGATCTCCCTGCTGAGAGAGCCGCCGCCAGCACTCTCCATGTCGTGAGAGAGAA 1274
 DB 311 AlaAspLeuProAlaGlnGlnProProGlnGlnGlnThrPheTyrGlnGlnProProLeu 330
 QY 1275 GAGAGGCTGTGTATGAGAGAACTTCCAGAGAGAGACCTTTCACGAGAGCCGCCACTG 1334
 DB 331 GlnGlnAlaValTyrGlnGlnProProGlnGlnGlnGlnThrPheTyrGlnGlnProProLeu 350
 QY 1335 GTGAGAGCAGAGGTGCTGCTGAGACCACTTGACACCAACATTCAGGGCCAGGGGCTC 1394
 DB 351 ValGlnGlnGlnGlnAlaGlySerGlnHisIleAspHisIleGlnGlnGlnGlnGln 370
 QY 1395 AGTGGGCAAGGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1454
 DB 371 SerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 390
 QY 1455 ATTCCTTTCAGCCGAGACCTTCATCAGGGCATGAGGTGATGAGAGAGGTGAGGTG 1514
 DB 391 IleSerPheAspProGlnAsnLeuIleThrGlyIleGlnValIleAspGlnGlyTyrPtp 410
 QY 1515 CGTGGCTATGGGCGGATGCGCATTTGGCATGTTCCTGCAACTACGTTGAGAGCTTAT 1574
 DB 411 ArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGlnLeuIle 430
 QY 1575 GAG 1577
 DB 431 Gln 431
 RESULT 5

QY 1455 ATCTCTTTGACCCCGAAGAACTTCATCAGGGGATCGAGGTGATCGACGAGGCTGTGG 1514
 DB 399 ILeSerPheAspProGluAsnLeuIleThrGlyIleGluValIleAspGluGlyTyrPTrp 418
 QY 1515 CGTGGCATGCGCGGATGGCCATTTTGCGATGTCCTCCCAACTACGTGAGCTCAT 1574
 DB 419 ArgGlyTyrClyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 438
 QY 1575 GAG 1577
 DB 439 Glu 439
 RESULT 6
 Q9AVH2 PRELIMINARY; PRT; 282 AA.
 AC Q9AVH2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative senescence-associated protein (Fragment).
 GN SSA-13.
 OS Pisum sativum (garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 CX NCBI_Taxid=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Ichihara wase; TISSUE=Pod;
 RA Parlaesa J.A.T.; Sunaga A.; Miyazaki T.; Hisaka H.; Sonoda M.,
 RA Nakagawa H.; Sato T.;
 RT "Cloning of cDNAs encoding senescence-associated genes, ACC synthase
 RT and ACC oxidase from stored snow pea pods (Pisum sativum L. var
 RT saccharatum) and their expression during pod storage.";
 RL Postharvest Biology and Technology 22:239-247(2001).
 DR EMBL: AB049723; BAB33421.1; -.
 FT NON_TER
 SQ SEQUENCE 282 AA; 31822 MW; FC29C7257D2D1059 CRC64;
 Alignment Scores:
 Pred. No.: 6,52e-48 Length: 282
 Score: 759.50 Matches: 187
 Percent Similarity: 68.28% Conservative: 11
 Best Local Similarity: 64.48% Mismatches: 67
 Query Match: 17.96% Indels: 26
 DB: 10 Gaps: 5
 US-10-028-952a-4 (1-2333) x Q9AVH2 (1-282)
 QY 838 GATCTGATTAAGGGGCTTCAGTCATATCCGACAGATGTAAGTTCGCCCATGGCT 779
 DB 11 AspergillusPenicilliumAspergillusAspergillusAspergillusAspergillus 30
 QY 778 CCGTCCGACGACATACACCAATGTCGTAACCTGGCTTCTCTGTAAGTGAAGGA 719
 DB 31 PheIntrProSerAlaMetThrAsnCysAlaAsnGlnArgPheLeuSerTyrTyrValGlu 50
 QY 718 TTACCATGGCAACAACATCATATAGTAAAGTAAATTAACCTGCTCCAGAGAGGCTCTAA 659
 DB 51 Leu--LeuLeuArgHisCysHisGlnTyrGlyValThrAsnLeuSerHisAspGlyLeuI 70
 QY 658 TCCGAGTCACGTTCCCTATATTAGTGGTGAACAATCCAGCTTGGAATTCGCTTCA 599
 DB 70 IepProAlaHisValPTrpTyrTyrAlaAsnAsnProThrIleuGlyGluPheCysPheT 90
 QY 598 CAATGATAGGAAGCGCGACATCCAGAGATCAAAAAGCAGCTGCTATGAACGCTTGGC 539
 DB 90 hrmelIleGlyArgAlaAspIleGluGlySerYsSerAsnValAlaMetAsnAlaTyrL 110
 QY 538 CGCCACAGACAGATTATCCCTGTGTGTAAGTTCGTAACCTGCTGCTTAAACCAAAA 479
 DB 110 eurProGlnAlaSerTyrProCysGlyAsnPheSerPheThrSerPheLys-PheArg 129

QY 478 GGTCAAGAAAGATCGTAGAGCCCGCTTTCACAGGTCTGTATTCTGACTGAAATCAAGTC 419
 DB 130 SerLeuYsAspArgLeuAlaThrLeuSerArgPheValPheValLeuGluIleArgIle 149
 QY 418 AAGCAGCTTTGACCTCTGCTCCACGAGAGTTCCTGCTCCCTGAGCTCGCC-TTA 360
 DB 150 LysArgAlaPheThrLeuLeuPheHisThrArgPheLeuPheSerLeuSerSerSerleu 169
 QY 359 GGACACTGCGCTTACCGTTTACAGAGTGTACCGCCCAAGTCAACTCCCACTGGCACT 300
 DB 170 GlyHisLeuArgTyrLeuLeuThrAspValProProGlnProAsnSerProAspAsn 189
 QY 299 GTCCCGGAGCGGGGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249
 DB 190 Val-----PheArgPro--AspArgProThrLysValSerLeuTrpIleL 204
 QY 248 AGAAGCGAGAGCCCTCGGGGCTGCGCCCGCGCTCAACGGGTGAGTGAAGAAACGATC 189
 DB 204 YsLysGluGlyGln-----ArgArgProProAspSerArgAsnLysLeuAsnValL 222
 QY 188 AGAGTAGTGTATTTCACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129
 DB 222 YsSerSerGlyIleSerThrPheAlaValSerSer-----SerHisLeuS 237
 QY 128 CGGGGACA-----CCGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
 DB 237 eTyrThrSerGlnValIleSerGlnSerArgThrArgValIleAsnArgValPhe- 256
 QY 80 TTCCCGGCTGATTCGCGCAAGCCGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 21
 DB 257 PheProAlaAspSerAlaLysAlaValProLeuLeuValSerLeuAspSerArgYs 276
 QY 20 GGACAGTGGCAATCTCGT 3
 DB 277 GlyGlnTyrGluPheArg 282
 RESULT 7
 Q80WPI PRELIMINARY; PRT; 436 AA.
 ID Q80WPI;
 AC Q80WPI;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Actin-binding protein 1.
 GN ABP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL6/J;
 RA Hou P.; Estrada L.; Gorecki J.L.;
 RT "Fgbl, a Cdc42-specific guanine nucleotide exchange factor, directly
 RT interacts with cortactin and Abp1 to modulate cell shape.";
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY098595; AAM28340.1; -.
 DR GO: GO:0005622; C:intracellular; IEA.
 DR GO: GO:0003779; F:actin binding; IEA.
 DR InterPro: IPR002108; Actbind_cofin.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00241; coflin_ADF_1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000066; SH3_1.
 DR SMART: SM00102; ADF_1.
 DR SMART: SM00326; SH3_1.
 DR PROSITE: PS00002; SH3_1.
 SQ SEQUENCE 436 AA; 48659 MW; 85AEF9781C698A3F CRC64;
 Alignment Scores:
 Pred. No.: 9.85e-44 Length: 436
 Score: 703.50 Matches: 152

Percent Similarity:	66.80%	Conservative:	15
Best Local Similarity:	60.80%	Mismatches:	22
Query Match:	16.63%	Indels:	62
DB:	11	Gaps:	5

US-10-028-952A-4 (1-2333) x Q80WP1 (1-436)

[illegible]

SEQUENCE FROM N.A.
TISSTSE=EMBRYO;
MEDLINE=98296439; PubMed=9630982;
Spartk A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
Nac. Biotechnol. 14:741-744(1996).
-1 SIMILARITY: CONTAINS A COPY OF THE SH3 DOMAIN.
CC -1 SIMILARITY: TO DREBRIN.
DR EMBL; U58884; AAC52640.1; -.
DR HSSP; P19174; 2HSF.
DR MGD; MG1.70C006; Dbnl.
DR GO; GO:0005737; C:cytosolasm; IDA.
DR GO; GO:0030027; C:lamilipodium; IDA.
DR GO; GO:0003779; P:actin binding; IDA.
DR GO; GO:0005315; F:protein binding; IDA.
DR InterPro; IPR02108; Actbind_coflin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofilin_ADF_1.
DR Pfam; PF00018; SH3_1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3_1.
DR SMART; SMO0102; ADF_1.
DR SMART; SMO0326; SH3_1.
DR PROSITE; PS50002; SH3_1.
KV SH3 domain.
FT DOMAIN 356 432 SH3.
SQ SEQUENCE 433 AA; 48428 MW; 602D3862C46FAAD CRC64;

Alignment Scores:
Pred. No.: 1.17e-43 Length: 433
Score: 702.50 Matches: 146
Percent Similarity: 72.65% Conservative: 16
Best Local Similarity: 65.47% Mismatches: 35
Query Match: 16.61% Indels: 26
DB: 11 Gaps: 5

US-10-028-952A-4 (1-2333) x Q62418 (1-433)

OY 971 GCCCGGCGCCGGAGAC-----CGGGGTCCGTGGGAGTGCCCTTCCTCGGAGAA 1021
Db 217 |A|A|A|A|A|G|A|G|A|G|A|G|A|G|A|G|T|G|T|G|I|n|G|I|n|G|I|n|H|s|A|G|S|E|R|A|A|G|I|A|A|P|R|O|S|E|R|g| 236
OY 1022 CGGGCGCGCGCGGAGAAAGCGCGGCCGCCCTCGCGCGTCAAGCAAGCACCGCACGCTTCGTGGG 1081
Db 237 Thrily-Glnproluginlu-----Alavalserrghtharginglutprgl 253
OY 1082 AACC-----TGCG 1099
Db 253 uSerAliglyngInAlaPrOHsPROATgLuIflePhelysgLnysGluaAgAlaMe 273
OY 1091 CTAAACACCCCTCATCTCCAAGTCTCTCAAGCTTCGCAAGGTGAAGAAGCCCCCTTCCTGCAGAA 1155
Db 273 tSetThrThrservaltnrserSergInPrGilylsLeuArgserPropheLeugInly 293
OY 1151 GCACCTACCCCAACAAGAGAACCCACTTTGGAGAGAGGCAAGCTGTCACCTCAAGGCC 1210
Db 293 sgInleuthrnInPrOGunnrserTYGLyArGsluPrOhrhAlarovalserghPr 313
OY 1211 CAGGCGAGATCCCTCGTGTGAAGAGCCGAGCCGACCACTCTCCATCTCTGTGACAGC 1277
Db 313 cAlAlaglyyal--CySGIugInuPrOlaPraoserThrrhuSer---AlaGIlnth 331
OY 1271 AGAAGAGAGGCTGTGTAAGAACCTCCAGAGAGAGAGACCTTCTACGAGAGAGCCCC 1333
Db 331 rGInGInuInuPrOhrThryrluvalAProPrOGInuInAsphrhleutyrgInuInuPr 351
OY 1331 ACTGGTCAGACGAGAGAGGTGCTGGCTCTGAGCAATTGACCAACCAACCAATTCAGGCGCAGG 1399
Db 351 OlenuValGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 371
OY 1391 GCTCAGTGGGAGAGAGGCTCTGTGCCTCGGAGCCCTTACACATACACAGGCGAGCGACGAC 1455
Db 371 yPheserclYlInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 391

QY 1451 AGAATCTCTTGAACCCCGAACCCTGATCGGCGATCGAGTGATCGACGAAGCTG 1510
 Db 391 TGLTLESERPhesapProgluanLeuilethnGlylleGluValIleaspGluGlytr 411
 QY 1511 GTGCGGTGCTATGAGCGCGATGCGCATTTTGCGATCTTCCCTGCACTAGCTGAGCT 1570
 Db 411 pTPARGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAntyrValGluLe 431
 QY 1571 CATTGAG 1577
 Db 431 uileGlu 433

RESULT 9
 Q8BH56 PRELIMINARY; PRT; 432 AA.
 AC Q8BH56; 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Dribin-like.
 GN DBNL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain, Eye, and Medulla oblongata;
 RX MEDLINE=22334683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., USCid T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Faley J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman U.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalska U., Smatius D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Mair M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Strausberg R.,
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK046073; BAC32592.1; -
 DR EMBL; AK053795; BAC35528.1; -
 DR EMBL; AK078082; BAC37118.1; -
 DR EMBL; BC046430; AAH46430.1; -
 DR MGI; MGI:700006; Dbnl.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:003027; C:lamellipodium; IDA.
 DR GO; GO:0003779; F:actin binding; IDA.
 DR GO; GO:0005515; F:protein binding; IDA.

DR GO; GO:0016601; P:RAC protein signal transduction; IDA.
 DR InterPro; IPR002108; Acctbind_cofin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; cofilin_ADF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00102; ADF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS0002; SH3; 1.
 DR SEQUENCE 432 AA; 48341 MW; 4A36D3B6E59C707F CRC64;
 SQ

Alignment Scores:
 Pred. No.: 4,58e-43 Length: 432
 Score: 694.50 Matches: 141
 Percent Similarity: 69.20% Conservative: 14
 Best Local Similarity: 62.95% Mismatches: 32
 Query Match: 16,428 Indels: 37
 DB: 11 Gaps: 5

US-10-028-952A-4 (1-2333) x Q8BH56 (1-432)
 QY 957 AGGCGCGTGGCCCGCGCGCGCGGACCGGGGTCCGGTGGAGTGCCCTTCGCTCG 1016
 Db 229 ArgSerAlaGlyAlaProArgThrGly----- 237
 QY 1017 GGAAACGGGCGCGCGCGGAAAGCGCGCGCCCTCGCCCTCAGCAGCAGCAGCTTCG 1076
 Db 238 -----GluProGluGluGlu-----AlaValSerArgThrArgGluGlu 250
 QY 1077 TGGGGAAC----- 1085
 Db 251 TPGLuSerAlaGlyGlnGlnAlaProHisProArgGluIlePheGlyGlnGluArg 270
 QY 1086 TGGCGCTAAACCACTCTCCTCAGTCTCAGCTCGGCGAGCTGAGAGCCCTTCCTG 1145
 Db 271 AlaMetSerThrThrSerValThrSerSerGlnProGlyLeuArgSerProPheLeu 290
 QY 1146 CAGAACACTCAACCAACCAAGACCCACTTGGCAGAGACCGAGCTGCTCCATCTCA 1205
 Db 291 GlnGlyGlnLeuThrGlnProGluThrSerTyrGlyArgGluProThrAlaProValSer 310
 QY 1206 AGGCCAGGCGAGATCTCCCTGCTGAGAGCGCGCGCCAGCACTCTCCATGCTGTG 1265
 Db 311 ArgProAlaAlaGlyVal---CysGluGluProAlaProSerThrLeuSer---Ala 328
 QY 1266 CAGGCAAGAGAGAGGCTGTATGAGAACTCCAGAGCAGAGAACTTTACAGAGCAG 1325
 Db 329 GlnThrGlnGlnGluProThrTyrGluValProProGluGlnAspThrLeuTyrGluGlu 348
 QY 1326 CCCCCCTGTCAGAGAGAGAGTGTGCTGAGCATTTGAGCACTTGACCAACCACTAGGCG 1385
 Db 349 ProProLeuValGlnGlnGlnAlaGlySerGlnHisIleAspAsnTyrMetGlnSer 368
 QY 1386 CAGGGCTAGTGGGCGAGGCTGTGTGCGCCGTGCTTACGACTTACAGGACGCCAGC 1445
 Db 369 GlnGlyPheSerGlyGlnGlyLeuGlnGlnAlaGlnAlaLeuTyrAspTyrGlnIleAlaAsp 388
 QY 1446 GACACAGATCTCTTGAACCCCGAGACCTCATCAGGCGATCGAGGTGATCGAGCA 1505
 Db 389 AspThrGlnIleSerPheAspProGluValLeuIlethnGlyIleGluValIleaspGlu 408
 QY 1506 GGCTGTGCGTGGCTATGAGCGCGAGATGCGCATTTTGGCATGTCCTTGCACACTACGTG 1565
 Db 409 GlyTyrTPArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrVal 428
 QY 1566 GAGCTCATGAG 1577
 Db 429 GluLeuIleGlu 432

RESULT 10
 Q9JHL4 PRELIMINARY; PRT; 436 AA.

```

AC Q9JHL4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE SH3P7R1.
GN SH3P7R1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB038365; BAA90867.1; -.
DR EMBL; AB039346; BAA90819.1; -.
DR HSSP; P19174; 2HSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPRO02108; Actbind_cofln.
DR InterPro; IPRO01452; SH3.
DR Pfam; PF00241; cofilin_ADP; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 436 AA; 48612 MW; 064CA1B6BAF84444 CRC64;

Alignment Scores:
Pred. No.: 6,5e-41 Length: 436
Score: 665,50 Matches: 138
Percent Similarity: 67,11% Conservative: 13
Best Local Similarity: 61,33% Mismatches: 35
Query Match: 15,73% Indels: 39
Gaps: 5

US-10-028-952a-4 (1-233) x Q9JHL4 (1-436)
QY 1014 CTGGGAAACGGGGCGCGCGGAAAGCGG-----CCG 1046
DB 214 LeuGInGluAlaAlaArgArgGluGlnArgTyrGlnGlnGlnHisArgSerAlaGlyPro 233
QY 1047 CCGCCCTGCGCCGTCAGCCACC-----GCACGTCGTGGGGA 1082
DB 234 ProSerProSerSerArgTnGlnGluLeuGlnGlnGlnValSerArgSerArgGln 253
QY 1083 ACCTGGCGC----- 1091
DB 254 GluTrpGluSerAlaGlnGlnGlnAlaProHisProArgGluLeuPheLysGlnLysGlu 273
QY 1092 -----TAAACACCTCCATCTCCAGTCTCAGCTGGCAAGTGAAGAGCCCTTC 1142
DB 274 ArgAlaMetSerTrpTrpSerValSerSerSerGlnProGlyLysLeuArgSerProPhe 293
QY 1143 CTGAGAGAGAGCGCTCACCAACGAGACCACTTGCGAGAGAGCGAGCGAGCTGTGCACATC 1202
DB 294 LeuGlnLysGlnPheTrpGlnProGlnAlaSerTyrGlnLysGlnProTrpSerProVal 313
QY 1203 TCAAGGCCCAAGGCGAGATCTCCCTGCTAGAGAGCGCGGCCAGCACTCTCCATGTCTG 1262
DB 314 SerArgProAlaAlaGlyVal---CysGlnGlnLeuAlaProSerTrpProSer--- 331
QY 1263 GTGAGGCGAGAGAGAGCGCTGTATGAGAACTCCAGAGCGAGAGCACTTCTACGAG 1322

```

```

DB 332 AlGlnThrAspAspGluProThrTyrGlnValProSerGlnGlnGlnThrLysGln 351
QY 1323 CAGCCCCACCTGATGAGACGAGCAAGGTGCTGCTGAGACACTTGACCACTTCAG 1382
DB 352 GluProProValGlnGlnProGlyAlaGlySerGlyHisLeuPheSerTrpMetGln 371
QY 1383 GCGCAGGAGGCTCAGTGGGCAAGGGCTGTGGCCCGGCGCTGTAGACACTACAGCAGCC 1442
DB 372 SerGlnAspLeuSerGlnGlnGlnLysLeuGlnAlaArgAlaLeuTrpAspTyrGlnAla 391
QY 1443 GACGACAGAGATCTCTCTTGAACCCGAGACCTCATCAGCGGCAATCGAGGTGATGAC 1502
DB 392 AspAspThrGlnLysSerPheAspProGluAsnLeuIleThrGlyIleGluValIleAsp 411
QY 1503 GAGCGTGTGGCGGTGCTATGGCGGAGATGGCACTTTGGCATGTCTCCCTGCACTAC 1562
DB 412 GluGlyTrpTrpArgGlyGlyGlyProAspGlyHisPheGlyMetPheProAlaIleTrp 431
QY 1563 GTGAGGCTCATTTGAG 1577
DB 432 ValGluLeuIleGlu 436

RESULT 11
Q9JMT4 PRELIMINARY; PRT; 429 AA.
ID Q9JMT4
AC Q9JMT4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE SH3P7r4 (fragment).
GN SH3P7R4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB038364; BAA90866.1; -.
DR HSSP; P19174; 2HSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPRO02108; Actbind_cofln.
DR InterPro; IPRO01452; SH3.
DR Pfam; PF00241; cofilin_ADP; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
FT NON TER
SQ SEQUENCE 429 AA; 47923 MW; A7E7E7BBAF4A1705 CRC64;

Alignment Scores:
Pred. No.: 7,06e-41 Length: 429
Score: 665,00 Matches: 136
Percent Similarity: 68,64% Conservative: 15
Best Local Similarity: 61,82% Mismatches: 35
Query Match: 15,72% Indels: 34
Gaps: 5

US-10-028-952a-4 (1-233) x Q9JMT4 (1-429)
QY 1014 CTGGGAAACGGGGCGCGCGGAAAGCGG-----CCGCCC 1049
DB 212 LeuGlnGlnAlaAlaAlaArgArgGlnGlnArgTyrGlnGlnGlnHisArgSerAlaGlyPro 231

```

```

QY 1050 CCTGCCCG-----TCACGACCCGACGTTCTGGGAGACC 1085
Db 232 ProArgThrGlyGluLeuGlnGlnValValSerArgSerArgGlnGlnTyrSerAla 251
QY 1086 TGGGCG-----TAAACC 1097
Db 252 GlyGlnGlnAlaProHisProArgGlnIlePheGlySerGlnGlnAlaGlnAlaMetSerThr 271
QY 1098 ACCGCACTCTCCAGTCTCTCAGCTGGCAGACTGAGAGAGCCCTTCTGACGAGAGAGCTC 1157
Db 272 ThSerValSerSerSerGlnProGlyLysLeuArgSerProPheLeuGlnGlnPhe 291
QY 1158 ACCCAACCAAGAGACCCATTGGCAGAGAGCCAGCTGCTCCATCTCAAGGCCAGGAGCA 1217
Db 292 ThrGlnProGlnAlaSerTyrGlyArgGlnProThrSerProValSerArgProAlaAla 311
QY 1218 GATGCTCCTGCTGAGAGCGCGGCGGACGACCTCTGCTGGTGAGGAGAGAG 1277
Db 312 GlyVal---CysGlnGlnIleValaProSerThrProProSer---AlaGlnThrAspAsp 329
QY 1278 GAGGCTGTATGAGAACCTTCCAGACGAGAGACCTTCTAGAGAGAGCCCGACCTGTG 1337
Db 330 GluProThrTyrGlnValaProSerGlnGlnIleThrLeuTyrGlnGlnProProVal 349
QY 1338 CAGGACGAGAGTGTGCTGCTGAGACATTCACACACATTCAGAGGCCGAGGGGCTCAGT 1397
Db 350 GlnGlnProGlyAlaGlySerGlyHisIleAspAsnTyrMetGlnSerGlnAspLeuSer 369
QY 1398 GGGGAGAGGCTGTGCGCCGCTGCTGCTGACACTACGAGAGAGCGGACGAGAGAGATC 1457
Db 370 GlyGlnGlnLeuGlnGlnAlaArgGlnAlaLeuTyrAspTyrGlnIleAlaAspAspThrGlnIle 389
QY 1458 TCCTTTGACCCCGAAGCTTCATACGCGCATCGAGTGTATCGACGAGAGCTGTGCTG 1517
Db 390 SerPheAspProGlnIleValaLeuIleThrGlyIleGlnValIleAspGlnGlyTyrTPPAsp 409
QY 1518 GGCATGAGGCGGAGATGGCATTTGGCATGTCCTCCCACTACGAGAGCTGATGAG 1577
Db 410 GlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGlnIleGln 429

RESULT 12
Q9Jm66 PRELIMINARY; PRT; 432 AA.
AC Q9Jm66;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SH3P7r3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shirao T.;
RT "Rattus rattus mRNA for SH3P7 (SH3P7r3).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB039819; BAA92709.1; -
DR HSBP; P19174; 2HSP
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofillin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR03452; SH3DOMAIN.
DR PRODOM; PPO00066; SH3; 1.
DR SMART; SMC0102; ADF; 1.
DR SMART; SMC0326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 432 AA; 48254 MW; E2D5E5C1D325B641 CRC64;

```

```

Alignment Scores:
Pred. No.: 1,286-40 Length: 432
Score: 661.50 Matches: 136
Percent Similarity: 72.09% Conservative: 19
Best Local Similarity: 63.26% Mismatches: 41
Query Match: 15.64% Indels: 19
DB: 11 Gaps: 6

US-10-028-952A-4 (1-2333) x Q9Jm66 (1-432)
QY 957 AGGCGGCGTCCCGCGCGCGCGGAGCGGGGTCGGAGTCCGAGTCCCTGCTCTG 1016
Db 229 ArgSerAlaGlyProProArgThrGlyGln-----LeuGlnGlnVal 243
QY 1017 GGAAACGGGCGCGCGCGGAA-----AGCGGCGCGCGCTCGCCG 1058
Db 244 ValSerArgSerArgGlnGlnIleThrGlnSerAlaGlnGlnAlaProHisProArgGln 263
QY 1058 -----TCACGACCCGACGTTCTGCGGGAACCTGGCGCTTAACCACTCCATCTCCAGT 1112
Db 264 IlePheGlnGlnGlnAlaArgAlaMetSer-----ThrThrSerValSerSer 279
QY 1113 CCTCAGCTGGGAGAGTGAAGAGAGCCCTTCTGACGAGAGCTCACCAACGAGAGACC 1172
Db 280 SerGlnProGlyLysLeuArgSerProPheLeuGlnGlnPheThrGlnProGlnAla 299
QY 1173 CACTTGGAGAGAGCCAGCTGCTGCTCCATCTCAAGGCCAGGAGATCTCCCTGCTAG 1232
Db 300 SerTyrGlyArgGlnProThrSerProValSerArgProAlaAlaGlyVal---CysGln 318
QY 1233 GAGCGGCGCGCGAGACCTCCATGTCGTGTCGACGAGAGAGAGAGAGAGCTGTATGAG 1292
Db 319 GlnLeuAlaProSerThrProProSer---AlaGlnThrAspAspGlnProThrTyrGln 337
QY 1293 GAACCTCCAGACGAGAGACCTTCTAGACGAGAGCCCGACCTGTCACAGAGAGTCT 1352
Db 338 ValProSerGlnGlnIleThrLeuTyrGlnGlnProProValGlnGlnProGlyAla 357
QY 1353 GGCCTGAGACATGACACACACATTCAGAGGCGCGGCTCAGTGGGCAAGGCTCTGT 1412
Db 358 GlySerGlyHisIleAspAsnTyrMetGlnSerGlnAspLeuSerGlyGlnGlyLeuGln 377
QY 1413 GCGCGTCCCTGTAGACTTACGAGCGACCGACGACGACGAGATCTCTTGAACCCGAG 1472
Db 378 AlaArgAlaLeuTyrAspTyrGlnIleAlaAspAspThrGlnIleSerPheAspProGln 397
QY 1473 AACCTCATCAGCGGATGAGGTGATCGACGAGAGCTGTGCGGCTATGAGCGCGAT 1532
Db 398 AsnLeuIleThrGlyIleGlnValIleAspGlnGlnIleTyrTPPAspGlyTyrGlyProAsp 417
QY 1533 GGCATTTGGCATGTCCTGCCAATCACTACGAGAGCTGATGAG 1577
Db 418 GlyHisPheGlyMetPheProAlaAsnTyrValGlnIleGln 429

RESULT 13
Q9Jm67 PRELIMINARY; PRT; 433 AA.
AC Q9Jm67;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SH3P7r2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shirao T.;
RT "Rattus rattus mRNA for SH3P7 (SH3P7r2).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

```

DR EMBL; AB039618; BA92708.1; -.
DR HSSP; P15174; 2HSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPRO02108; Actbind_cofin.
DR InterPro; IPRO01452; SH3.
DR Pfam; PF00241; Cofilin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR Prodom; PD000065; SH3; 1.
DR SMART; SMO0102; ADF; 1.
DR SMART; SMO0326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain.
KW SEQUENCE 433 AA; 48341 MW; 952FFB7E186AEDDA CRC64;

Alignment Scores:	
Pred. No.:	1,4e+0
Score:	66.00
Percent Similarity:	68.02%
Best Local Similarity:	61.26%
Query Match:	15.63%
AB:	11
Length:	433
Matches:	126
Mismatch:	35
Mismatches:	35
Indels:	5
Gaps:	5

US-10-028-952A-4 (1-2333) x Q9JM67 (1-433)

QY	1014	CTGGGAAAAGGGGGCGGGCGGAAAGCGG-----CCG	1046
Db	214	LeuIngluaLaLaArgaYrgLuglnatYrgIngluInhiSaYserLaAgiPro	233
QY	1047	CCCCCTCGCCG-----TCAGCAACCGCAGCTTCGTGGGA	1082
Db	234	ProSerArghThrgLgYluLeuIngluIngluValValSerArghSerArYgIngluIntrGlu	253
QY	1083	ACC-----TGCGCG	1091
Db	254	SerLaagiYgIngluaLaProhiLaProArYgLuLePheLysGlnYsGluArYgLaMet	273
QY	1092	TAAACCAACCTCATCTCCAGTCCCTCAGCCTGGCAGCTGAGAGAGACCCCTCTCTGACAGAG	1151
Db	274	SerThrhSerValSerSerSerGlnProLiyLysLeuArghSerProPheLeuGlnLys	293
QY	1152	CAGCTCACCCACCAAGAGAGCCCACTTTGGCAGAGAGCCAGCTGCTGCCATCTCAAGGCC	1211
Db	294	GlnPheThrgInProGluLaSerTyrGlyArYgInuProThrhSerProValSerArYgPro	313
QY	1212	AGGGGAGATCTCCCTGCTGAGAGAGCGGGCGCCAGCACTCCTCATGCTGTGTACAGCA	1271
Db	314	AlaLaagiLysLa---CysGlnGluLeuLaProSerThrhProProSer---AlaGlnThr	331
QY	1272	GAGAGAGAGAGCTGTGTATGAGGAACTTCAGAGCAGAGAGACCTTCTACAGAGACCCCCA	1331
Db	332	AspAspGlnuProThrhTyrGluValProSerGlnGlnGluThrhLeuTyrGlnGluInuPro	351
QY	1332	CTGTGTGCAGCAGCAAGAGTGTGTGCTCTGAGACATGTGCACCAACCATTCAGAGGCGAGGG	1391
Db	352	ProValGlnGlnProGluYLaagiYserGlyHisLeaPaanTyrMetGlnSerGlnhsp	371
QY	1392	CTCAGTGGGCAAGGGCTGTGTGCCCGTGCCTGTACGACTACAGGCGAGCCGAGACACA	1451
Db	372	LeuSerGlyGlnGlyLeuCyLehLaArgLaLeuTyrAspTyrGlnLaLaLaAspAspThr	391
QY	1452	GAGATCTCTCTTGACCCCGAGAACTCATCACGGGCACTCGAGGTATGAGAGAGCGTGG	1511
Db	392	GluLeSerPheAspProGluLysnLeuLeThrhGlyIleGluValIleAspGlnGlyTyr	411
QY	1512	TGGCGAGGTTATGGGCGCGATGACCATTTGGCATGTTCCCTGCACCACTACGTGAGCTC	1571u
Db	412	ThraYgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeu	431
QY	1572	ATTGAG	1577
Db	432	IleGlu	433

RESULT 14

Q7ZXQ9
ID Q7ZXQ9 PRELIMINARY; PRT; 447 AA.

AC 01-JUN-2003 (TREMBLER). 24, Created)
 DT 01-JUN-2003 (TREMBLER). 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLER). 25, Last annotation update)
 DE Similar to drebrin-like.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
 OC Xenopodidae; Xenopus.
 RX NCBI_TaxID=8355;
 RN [1]
 RP
 RP TISSUE=Embryo;
 RA Klein S., Strausberg R., to the EMBL/GenBank/DBJ databases.
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC044296; AAH44296.1;
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003779; F:actin binding; IEA.
 DR InterPro; IPR02108; Acbhd_cofin.
 DR InterPro; IPR01452; SH3.
 DR Pfam; PF00241; cofilin_ADP; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PRO0452; SH3DOM1N.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00102; ADP; 1.
 DR SMART; SW00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 SQ SEQUENCE 447 AA; 50611 MW; 3BC6767EE63B521 CRC64;

Alignment Scores:	
Pred. No.:	1,17e-20
Score:	392.50
Percent Similarity:	59.32%
Best Local Similarity:	48.59%
Query Match:	9.48%
DB:	13
Gaps:	2
Length:	447
Matches:	66
Conservative:	15
Mismatches:	49
Indels:	23

US-10-028-952A-4 (1-2333) x Q7ZXQ9 (1-447)

```

QY      1110 AGTCCTCAGCTGGCAAGCTGAGAGGCCCTCTCTGCAAGAGCAGTCAACCCAAACAGAG 1169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      273 SerAlaGlnProGluYArgLeuAlaSerProHeuLeuGlnLysSerAlaCyGlnProGlu 292
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1170 ACC-----CACTTGGCAGAGAGCCAGCTGCTGCATCTCAAG 1208
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      293 SerSerProProSerProValHisArgValGlnGluProProSerProProValTyr 312
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1209 CCCAGGGCAGATCTCCCTGCTGAGAGCGCGCGCC-----AGACATCTTCA----- 1256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      313 ProAlaHisGlnThrProProGluSerProValProProValSerHisProProGluSer 332
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1257 -----TGTCTGGTGT---CAGGCAGAGAGAGAGAGCTGTGATAG 1292
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      333 ThrValHisValLysGluGlnCysThrLysSerGlnGlnGlnGlnGlnLysGln 352
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1293 GAACTCCAGAGCAGAGACCTTTACAGAGAGCGCCCACTGCTGTGACAG-----CAA 1344
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      353 AspAlaThrGluAspGluAsnLysThrGlnAspThrThrGluAsnGlnAsnLysThrGlu 372
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1347 GGTCGTGGCTGAGACATTCAGACCAACCACTTACAGGGCCAGGGGCTCAGTGGCCAGAG 1406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      373 AspThrProGlnGlnGluProValTyrGluLysGluValGlu-----GluGluLysGly 390
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1407 CMTGTGGCCCGGACCTGTACAGTACAGGAGCGGACGACACAGAGATCTCTTTCAG 1466
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      391 ValCysAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspThrGluLysSerHeAsp 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1467 CCCGGAACCTCATCAGCGGACATGAGGTATGACGAAGGCTGTGGCGGTATATGG 1526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 411 ProbsapleuilethrginileglnPheileAspGluGlyTrpTrpArgglyPheSer 430

QY 1527 CCGGATGGCCATTTGGCATGTTCCCTGCCACACTAGTGGAGCTCATTTAG 1577

Db 431 ProaladiylisPheglyMetPheProAlaAsnTyrValGluLeuLeuGlu 447

RESULT 15

Q9XUT0 PRELIMINARY; PRT; 643 AA.

AC Q9XUT0; 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE K08E3.4 protein.

GN K08E3.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_Taxid=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA McMurray A.A.;

RL Submitted (NCV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology."

RT Science 282:2012-2018(1998).

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

DR EMBL; Z81568; CAB04592.1; -

DR PIR; T23453; T23453.

DR HSP; P29355; ISEM.

DR McompPep; K08E3.4; CE18868.

DR GO; GO:0005622; C:intracellular; IEA.

DR GO; GO:0003779; F:actin binding; IEA.

DR InterPro; IPR002108; Actbind_cofin.

DR InterPro; IPR01452; SH3.

DR Pfam; PF00241; cofilin_ADF; 1.

DR PRINTS; PR00018; SH3; 1.

DR PRINTS; PR00452; SH3DOMAIN.

DR PRODOM; PD000066; SH3; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50002; SH3; 1.

DR SH3 domain.

QW SEQUENCE 643 AA; 71941 MW; B95BP648425EP694 CRC64;

Alignment Scores:

Pred. No.: 6,35e-13 Length: 643

Score: 288.50 Matches: 92

Percent Similarity: 38.04% Conservative: 32

Best Local Similarity: 28.22% Mismatches: 105

Query Match: 6.82% Indels: 97

DB: 5 Caps: 11

US-10-028-952a-4 (1-2333) x Q9XUT0 (1-643)

QY 841 GCCCAGGCGGAAACGATACGCGAGCGCGGAGCCTCGGTGCTCGGATACCGCGATCC 900

Db 335 AlaglnlysglnysglnupheglutThrAspArglyLeuAlaAspAspLeuHlsGly 354

QY 901 CCGGCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGAG 960

Db 355 LysAlaGlnIleSerAspLysProValProSerSerValAsnProAla----- 370

QY 961 CGCGTGGCGGAA 1020

Db 371 -----ProAlaProAlaProThrProSerAlaGlyLeu---ValGlySerArgGlyGlu 387

QY 1021 ACGGGGGGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1080

Db 388 LeuPheSerSerLysProSerGlyProValLeuProLys-----ProGlnValaangly 405

QY 1081 GAACCTGGCGGCTAAACCACTTCATCTTCAGACTCCTCAGCCTTGGGAGAGAGCCCT 1140

Db 406 SerProLysLysTrpProPro--ValGlyThrThrSerProPro-----ArgGluPro 423

QY 1141 TCCTCGAGAGAGAGCTCACCACAGAG----- 1179

Db 423 al-----AsnArgLeuThrGluProAspGluProSerThrThrThrProLysProIle 441

QY 1180 CGAGAGACCCAGCTGCTGCCATCTCAGAGCCAGGCGAGATCTCCCTGCT----- 1229

Db 441 lATyGluProGluProMetValTyrLysProGluAlaMetLysProAlaValSerTyr 461

QY 1230 -----GAGAGCGCGGCGCCAGACT----- 1250

Db 461 sPAlaTyrGluGluProProAlaGluProAlaProProThrPheLeuAlaProThrPro 481

QY 1251 -----CCTCCATGCTGTGTGACGAGAAAGAGAGGCTGTATG 1290

Db 481 alIleAlaProProProProGluProThrProAlaProSerHisTyrAlaSerGlnTyr 501

QY 1291 AGGAACCTCCAGAGAGAGACCTTC----- 1316

Db 501 sPAlaProProValHisGlnSerPheGluProAlaValProProValSerAlaProSerH 521

QY 1317 -----TACGAGAGCCCGCACTGTGTGACAGACAGAGAGTGTGCTGTGAGC 1362

Db 521 lATyrAlaSerGlnTyrAspAlaProProGluProIleAspSerHisSerSerSerG 541

QY 1363 ACATTGACCAACCACTTAGGCGCAG----- 1388

Db 541 lnteuProAlaHisIleAlaSerGlnTyrAspMetProProValMetProGluGluPro 561

QY 1388 ----- 1388

Db 561 alPheAlaProLysSerSerProIleLysValAlaAlaProProIleAspGlnTyrAsp 581

QY 1389 -----GGGCTCAGTGGGAGAGGCTCTGTGCCCGCTGTACACTTACAGGAG 1440

Db 581 heProProAlaValAlaGluGlnAsn-----AlaMetAlaLeuTrpAspTyrGlnAla 599

QY 1441 CCGAGACAGACAGATCTCTTTGACCCCGAGAACTCATCAGGAGCATCAGGATGATCG 1500

Db 599 lAspAspThrGlnIleSerPheAspProAlaAspIleIleThrAspIleAspGlnVal 619

QY 1501 ACGAAGCTGTGCGGTAGTGGCGGATGGCCATTTGGCATGTCTCCCTGCCAACT 1560

Db 619 sPserGlyTrpTrpLysGlyArgAlaProSerGlyArgValGlyLeuPheProAlaAsn 639

QY 1561 ACGTGAGCTCAT 1574

Db 639 YrValLysLeuIle 643

Search completed: April 1, 2004, 15:58:20

Job time : 139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 1, 2004, 15:47:06 ; Search time 25 Seconds
(without alignments)
9718.374 Million cell updates/sec

Title: US-10-028-952a-4
Perfect score: 4230
Sequence: 1 gcacgagatccaccatcgtcc.....agcttgactcttcgcga 2333

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool.P/US10028952/runat.01042004.154625.643/bpp.query.fasta.1.2503
-DB=SwissProt 42 -QFWT=fastan -SUFFIX=rsf -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdt -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pct -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USFR=US10028952.@CGC.1.1.25@runat.01042004.154625.643 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOC=100 -XON2LOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	6.5	486	1 HSI_HUMAN	P14317 homo sapien
2	261.5	6.2	486	1 HSI_MOUSE	P49710 mus musculu
3	244.5	5.8	563	1 SRCB_CHICK	Q01406 gallus gall
4	242.5	5.7	660	1 YHL1_EBV	P03181 epstein-bar
5	231	5.5	546	1 SRCB_MOUSE	Q60598 mus musculu
6	230	5.4	550	1 SRCB_HUMAN	Q14247 homo sapien
7	214	5.1	857	1 APP1_SCHPO	O9P768 schizosacch
8	195.5	4.6	5262	1 MLT2_HUMAN	O14656 homo sapien
9	193.5	4.6	825	1 SES_RAT	O63003 rattus norv
10	191.5	4.5	660	1 YHL1_EBV	P03181 epstein-bar
11	191	4.5	775	1 ICP0_HSV11	P02457 herpes simp
12	189	4.5	1455	1 CA11_CHICK	P02457 gallus gall
13	186.5	4.4	1355	1 CA21_RANCA	O42350 rana catesb
14	185.5	4.4	1183	1 DRPL_RAT	P54258 rattus norv
15	184.5	4.4	1336	1 W146_HUMAN	P54258 rattus norv
16	184.5	4.4	1453	1 CA11_MOUSE	P11087 mus musculu
17	182	4.3	592	1 ABP1_YEAST	P15891 saccharomyc
18	181.5	4.3	1355	1 CA21_RANCA	O42350 rana catesb

19	180.5	4.3	1426	1 BCL9_HUMAN	O0512 homo sapien
20	180.5	4.3	1460	1 CA11_CANFA	O9x617 canis famli
21	180	4.3	617	1 ABP1_SACEX	P38479 saccharomyc
22	178.5	4.2	1262	1 CA13_CHICK	P12105 gallus gall
23	178.5	4.2	1372	1 CA21_MOUSE	Q01148 mus musculu
24	178	4.2	699	1 VGLG_HSV2H	F13290 herpes simp
25	177.5	4.2	1464	1 CA13_MOUSE	P08121 mus musculu
26	175.5	4.2	1466	1 CA13_HUMAN	P02461 homo sapien
27	175.5	4.1	2716	1 OSA_DROME	Q81n94 drosophila
28	175	4.1	3530	1 MY15_HUMAN	Q9ukn7 homo sapien
29	174.5	4.1	1003	1 MED6_HUMAN	Q96d66 homo sapien
30	174	4.1	1185	1 DRPL_HUMAN	P54258 homo sapien
31	173.5	4.1	1461	1 LE18_PRVIF	P11675 pseudorabie
32	172.5	4.1	1464	1 CA11_HUMAN	P02451 homo sapien
33	172	4.1	439	1 XP2_XENLA	P17437 xenopus lae
34	172	4.1	1459	1 CA12_MOUSE	P28481 mus musculu
35	172	4.1	1321	1 TRS2_MOUSE	P31123 pseudorabie
36	171.5	4.1	1321	1 SHK1_HUMAN	Q9Y566 homo sapien
37	171.5	4.1	2161	1 SHK1_HUMAN	Q9umt6 homo sapien
38	171.5	4.1	2715	1 MLT4_HSV2H	P28284 herpes simp
39	171	4.0	825	1 CA21_HUMAN	P08123 homo sapien
40	170.5	4.0	1366	1 CA21_BOVIN	P02453 bos taurus
41	170	4.0	779	1 ICP0_HSV2H	P28284 herpes simp
42	170	4.0	825	1 ICP0_HSV2H	O46392 canis famli
43	170	4.0	2716	1 CA21_CANFA	O81n94 drosophila
44	170	4.0	2716	1 OSA_DROME	Q96d66 homo sapien
45	168.5	4.0	1003	1 MED6_HUMAN	

ALIGNMENTS

RESULT 1	ID	HSI_HUMAN	STANDARD	PRT	486 AA.
AC	P14317				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	15-MAR-2004 (Rel. 43, Last annotation update)				
DE	Hemopoietic lineage cell specific protein (hematopoietic cell-specific LTN substrate 1) (LCKBP1).				
DE	Hemopoietic lineage cell specific protein (hematopoietic cell-specific LTN substrate 1) (LCKBP1).				
GN	HLS1 OR HSI.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lymph;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,				
RT	"Isolation and characterization of a novel human gene expressed specifically in the cells of hematopoietic lineage."				
RL	Nucleic Acids Res. 17:9367-9379(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lymph;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,				
RT	"Isolation and characterization of a novel human gene expressed specifically in the cells of hematopoietic lineage."				
RL	Nucleic Acids Res. 17:9367-9379(1989).				

RT Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002) .
RN [3]
RP SEQUENCE OF 97-108; 193-201 AND 240-248.
RX MEDLINE=96311348; PubMed=8713105;
RA Egeron M., Moritz R.L., Draker B., Kelso A., Simpson R.J.;
RT "identification of the 70kd heat shock cognate protein (Hsc70) and
RT alpha-actinin-1 as novel phosphotyrosine-containing proteins in T
RT lymphocytes." ;
RL Biochem. Biophys. Res. Commun. 224:666-674(1996) .
RN [4]
RX BINDING TO HAX-1 PROTEIN.
RX MEDLINE=97211841; PubMed=9058808;
RA Suzuki Y., Demoliere C., Kitamura D., Takeshita H., Deuschle U.,
RA Matanabe T.;
RT "HAX-1, a novel intracellular protein, localized on mitochondria,
RT directly associates with Hs1, a substrate of Src family tyrosine
RT kinases." ;
RT J. Immunol. 158:2736-2744(1997) .
RN [5]
RP PHOSPHORYLATION OF TYR-222.
RX MEDLINE=99167537; PubMed=10066823;
RA Brunetti A.M., Donella-Deana A., James P., Quadroni M., Contri A.,
RA Marin O., Pina L.A.;
RT "molecular features underlying the sequential phosphorylation of Hs1
RT protein and its association with c-Fgr protein-tyrosine kinases". ;
RT J. Biol. Chem. 274:7557-7564(1999) .
CC -1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine
CC kinase. Plays a role in antigen receptor signaling for both
CC clonal expansion and deletion in lymphoid cells. Directly
CC associates with HAX-1, through binding to its C-terminal region.
CC May also be involved in the regulation of gene expression.
CC -1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.
CC -1- SUBCELLULAR LOCATION: Mitochondrial (Probable) .
CC -1- TISSUE SPECIFICITY: Expressed only in tissues and cells of
CC hematopoietic origin.
CC -1- DEVELOPMENTAL STAGE: Expressed in early stage of myeloid and
CC erythroid differentiation.
CC -1- PFM: Phosphorylated by LYN, rapidly after crosslinking of surface
CC Igm on B cells.
CC -1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN) .
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16663; CA34651.1; -
DR EMBL; BC016758; AAH16758.1; -
DR F1R; 507633; S07633.
DR HSSP; P29335; ISEM.
DR Genew; HGNC:4844; HCLSL.
DR M1W; 601306; -
DR GO; GO:0005655; C:DNA-directed RNA polymerase II, core complex; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007242; P:intracellular signaling cascade; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IP003138; Hs1/Cortactin.
DR InterPro; IP000108; Neu_cyc_fac2.
DR InterPro; IP001452; SH3.
DR Pfam; PF02218; Hs1_rep; 4.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; P00499; P67PHOX.
DR PRINTS; P00452; SH3DOMAIN.
DR Prodom; PD000066; SH3; 1.
DR SMART; SMD0326; SH3; 1.
DR PROSITE; PS5002; SH3; 1.
KW Repeat; SH3 domain; Phosphorylation.

FT	DOMAIN	27	66	INVOLVED IN HAX-1 BINDING.
FT	DOMAIN	81	214	3.5 X 37 AA TANDEM REPEATS.
FT	REPEAT	81	116	1.
FT	REPEAT	117	153	2.
FT	REPEAT	154	190	3.
FT	REPEAT	191	214	4 (INCOMPLETE).
FT	DOMAIN	428	486	SH3
FT	MOD_RES	222	222	PHOSPHORYLATION.
FT	CONFLICT	241	242	KF -> FK (IN REF. 3).
FT	SEQUENCE	486 AA;	53998 MW;	61AE63715DF5DF2 CRC64;

Alignment Scores:

Pred. No.:	1.57e-08	Length:	486
Score:	277.00	Matches:	87
Percent Similarity:	42.97%	Conservative:	23
Best Local Similarity:	33.98%	Mismatches:	85
Query Match:	6.55%	Indels:	61
	1	Gaps:	10

US-10-028-952A-4 (1-2333) x HSL_HUMAN (1-486)

QY	938	CGCGCGCCCGCGCGCGCGCGCGCTGCCCCCGCGCGCGCGG-----GACCGCGGCT	991
DB	250	ArglysaRgsluglulguluglulylalaglnlvalalargatrgslnglulgyls	269
QY	992	CCGGTG-----CGGAGTGCCTTCCTGCTCGGGAACGGGGCGGGCGGGAAGCGGGC	1044
DB	270	AlaValThrlrlysaRserProgluAlaProgluProvallllealmerglugluProla	289
QY	1046	GCCCCCTGCGCGCGCGCGCGCGCGCGCTGCGGGAACCTGCGGCTAAACCACTTCAT	1108
DB	290	ValProAlaProleuProlyllys-----	297
QY	1106	CTCCAGTCTCAAGCTCTGCGCAAGCTGAGAGGCCCTTCCTGAGAAAGCAAGCTCAACCAAC	1166
DB	298	llesersRgluAlatrpProProvalglYThrProProserSergluSergluProval	317
QY	1166	AGAGACCACCTTGTGGCAGAGAGCAGCTGTCCATCTCAAGGCCAGGAGCATCTCC	1222
DB	318	Arg-Thr-----SerArglunHlsProvalProleuProIleArglnThrleuPr	335
QY	1226	TGCT-----GAGAGCGCGCGCGCGCGCGCGCTCCATGCTCG-----GTGCAGGC	1276
DB	335	CGluhspsnslugluPro-----ProAlaLeuProProRgrThrlreuglulgylleughlva	354
QY	1271	AGAGAGGAGGCGCTGTATGAGAACCTTCAGAGCAGGAGACCTTCTACGAGCAGCCCC	1333
DB	354	lgluglulgluProvalYrgluAlagluProgluProgluPro-----gluProgluPr	372
QY	1331	ACTGCTGAGAGCAGCAGAGGTGCTGCTGAGACACTTACACCAACCATTCAG-----	1387
DB	372	CGluProgluhsnpsRlyrgluAspvalglugluMetcspRgrHsglulglnluAspgl	392
QY	1382	-----	1387
DB	382	uProgluGllyAspYrgluGlulvalleugluProgluAspserSerPheSerAlaIe	412
QY	1383	-----GGCAGCGGCGCTCAAGTGGCAGCGGCTCTG	1411
DB	412	uAlaIyserSerRglYcysProAlaglyAlaIaglyAlaIaglyAlaIaleuglyllese	432
QY	1412	TGCCCGTCCCTGACGACTACACGAGCGCAGCAGCAGAGATCTCTTTGACCCGCA	1477
DB	432	rAlaValAlaLeuYrAspYrGlnglYgluYserAspRgluAspSerPheAspProhs	452
QY	1472	GAACTCATACGCGGCATCGAGGTGAGCAGCAAGAGCTGTGCGCTGTATGAGCCGGA	1533
DB	452	pAspVallelthrAspIlegluMetValAspIuglylrPrtpRgYly---ArgCyshl	471
QY	1532	TGGCCATTTTGGCAGTTCCTCGCCCACTCACTCGAGAGCTCATGAG	1577
DB	471	sglylHspRglYleuPheProAlaAsmYrAllylYsleuLeuglu	486

RESULT 2			
ID	HS1_MOUSE	STANDARD:	PRT: 486 AA.
AC	P49710		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Hematopoietic lineage cell specific protein (Hematopoietic cell-specific LYN substrate 1) (LCKBPL).		
CN	HCUS1 OR HSL.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=95217198; PubMed=7535527;		
RA	Kitamura D., Kaneko H., Taniuchi I., Yamamura K.I., Watanabe T.;		
RT	"Molecular cloning and characterization of mouse HSL.";		
RL	Biochem. Biophys. Res. Commun. 208:1137-1146(1995).		
RP	[2]		
RC	SEQUENCE FROM N.A.		
FX	STRAIN=BA1B/C;		
RA	MEDLINE=97098949; PubMed=8943564;		
RT	Takemoto Y., Sato M., Furuta M., Hashimoto Y.;		
RL	"Distinct binding patterns of HSL to the Src SH2 and SH3 domains reflect possible mechanisms of recruitment and activation of downstream molecules.";		
CC	Int. Immunol. 8:1699-1705(1996).		
CC	-1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine kinase. Plays a role in antigen receptor signaling for both clonal expansion and deletion in lymphoid cells. Directly associates with HAX-1, through binding to its C-terminal region. May also be involved in the regulation of gene expression (By similarity).		
CC	-1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.		
CC	-1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).		
CC	-1- TISSUE SPECIFICITY: Expressed only in tissues and cells of hematopoietic origin.		
CC	-1- PTM: Phosphorylated by LYN; rapidly after crosslinking of surface Igm on B cells (By similarity).		
CC	-1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).		
CC	-1- SIMILARITY: Contains 1 SH3 domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on ways that use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; D42120; BAA07701.1; -		
DR	EMBL; X84797; CAA59265.1; -		
DR	PIR; I49760; I49760.		
DR	HSSP; P29355; ISEM.		
DR	MGD; MG1:104568; HCLSL.		
DR	GO; GO:0017124; F:SH3-domain binding; IDA.		
DR	InterPro; IPR000314; Hsl/Cortactin.		
DR	InterPro; IPR001452; SH3.		
DR	Pfam; PF02218; Hsl_rep; 4.		
DR	Pfam; PF00018; SH3; 1.		
DR	PRINTS; PR00459; P67PHOX.		
DR	PRINTS; PR00452; SH3DOMAIN.		
DR	PRODom; PD000066; SH3; 1.		
DR	SMART; SM00326; SH3; 1.		
DR	PROSITE; PS50002; SH3; 1.		
KW	Repeat; SH3 domain; Phosphorylation.		
FT	DOMAIN 27 66 INVOLVED IN HAX-1 BINDING (BY SIMILARITY).		
FT	DOMAIN 81 214 3_5 x 37 AA TANDDEM REPEATS.		
FT	REPEAT 81 116 1.		
FT	REPEAT 117 153 2.		

[illegible]

ID	SRCS_MOUSE	STANDARD	PRT	546 AA.
Db	468	-----	468	-----
Qy	590	GGAGAGCCGACACATCGAAGATCAAAACGACGCTGATGAAAGCTTGCCGCCCA	531	-----
Db	469	GlyProAlaasp	ProProAla	475
Qy	530	GCACGATATCCCTGTGTGTAACCTTTCTGACACCTCTCGTTAAACCCAAAGCTGAA	471	-----
Db	476	AlaAlaArg	LeuProProGluArgGln	485
Qy	470	GGATCGAGAGCCCGCTTCAACGCTGTATCTGACTGAAATTCAGATCAAGCGAGC	411	-----
Db	486	-----	ProArgLeuPro	494
Qy	410	TTT-----TGCCCTTCTGCTCCACGGAGATTCTGTCTCCCGAAGCTCGCTTAGA	357	-----
Db	495	AlaGlnArgysProAlaGlyPro	ProProThrArgSerGly	508
Qy	356	CACCTGCGTTACCGCTTTGACAGGTGACCG-----CCCACTCAACTCCCACT	306	-----
Db	509	AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAspPro	528	-----
Qy	305	GGACATCTCCCGGAGGGGTGCGCCCGCGCGCGCGG-----CCGG	261	-----
Db	529	Gly-----CysProAlaGlnThrTPAArgArgSerGlyAlaGlnArgGlyHisProProG	548	-----
Qy	260	CGCGTTGGCCGACGAAGACGACGCCCTCGGAGCTCGGCC-----	220	-----
Db	548	lyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProG	568	-----
Qy	219	-----CCCGCTCACCGGGGTGAGTAAAAAGATGACG---AGTAGGTGTTTACCGG	168	-----
Db	568	lyThrProAlaAlaAlaProGlyProGlyGlyAlaAlaAlaAlaProSerGlyAlaThrProH	588	-----
Qy	167	CGCGCCGACGAGGGCGGAGACCCCGCGGCGCCCTCGCGGG---ACACCGGAGGAGGC	111	-----
Db	588	ISPProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluA	608	-----
Qy	110	GCOCGGGGGCCCAAGCTC	94	-----
Db	608	rgGngGluProArgLeu	613	-----

RESULT 5

ID	SRCS_MOUSE	STANDARD	PRT	546 AA.
AC	06058:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Src substrate cactactin.			
GN	EMSI OR CTN.			
OS	Mus musculus (Mouse).			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
CK	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALE/c;			
RX	MEDLINE=94268839; Pubmed=751662;			
RA	Michlajese M.R., Mannion-Henderson J., Wu H., Parsons J.T.,			
RT	Bender T.P.,			
RT	"The protein tyrosine kinase substrate cactactin is differentially			
RT	expressed in murine B lymphoid tumors.";			
RL	Oncogene 9:1989-1997(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 125-138; 273-289 AND 534-543.			
RX	MEDLINE=9404284; Pubmed=7693700;			
RA	Zhan X., Hu X., Hampton B., Burgess W.H., Priesel R., Maciag T.;			
RT	"Marine cactactin is phosphorylated in response to fibroblast growth			
RT	factor-1, on tyrosine residues late in the G1 phase of the BALB/c 3T3			
RT	cell cycle.";			
RL	J. Biol. Chem. 268:24427-24431(1993).			

```

CC -1- FUNCTION: May contribute to the organization of cell structure.
CC The SH3 motif may function as a binding region to cytoskeleton.
CC Tyrosine phosphorylation in transformed cells may contribute to
CC cellular growth regulation and transformation.
CC -1- SUBUNIT: Interacts with SHANK2 and SHANK3 via its SH3 domain (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Detected in most murine tissues, but not
CC detected in B lymphocytes or plasma cells.
CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U03184; AAA19689.1; -.
CC DR PIR: I48899; I48899.
CC DR HSSP: P19174; ZHSP.
CC DR MGD: MGI:99695; Cctn.
CC DR InterPro: IPR003134; Hsl/Contractin.
CC DR InterPro: IPR001018; Neu_cyt_fact_2.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF02218; Hsl_rep; 7.
CC DR Pfam: PF00018; SH3_1.
CC DR PRINTS: PR00499; P67PROX.
CC DR PRINTS: PR00452; SH3DOMAIN.
CC DR PRODOM: PD000066; SH3; 1.
CC DR SMART: SMO0326; SH3; 1.
CC DR PROSITE: PSS0002; SH3; 1.
CC KW Phosphorylation; Repeat; SH3 domain; Cytoskeleton.
CC MW 83
CC FT REPEAT 83 329 6.5 x 37 AA TANDEN REPEATS.
CC FT REPEAT 83 119 1.
CC FT REPEAT 120 156 2.
CC FT REPEAT 157 193 3.
CC FT REPEAT 194 230 4.
CC FT REPEAT 231 267 5.
CC FT REPEAT 268 304 6.
CC FT REPEAT 305 329 7 (INCOMPLETE).
CC FT DOMAIN 488 546 SH3.
CC FT COMFLICT 9 9 A -> R (IN REF. 2).
CC SQ SEQUENCE 546 AA; 61260 MW; 8F93A026ADD1D6DAF CRC64;

```



```

RESULT 7
APPL_SCHPO STANDARD; PRT: 857 AA.
ID APPL_SCHPO Q979EB; P78815;
AC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE APPL protein.
GN APPL OR SPAPJ760.02C
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_Taxid=4896;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Galliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby I., Howarth S., Huckle E.J., Hunt S., Jagsle K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkert G., Art R., Robben J., Grynolopez B.,
RA Welford J., Vansteede E., Rieger M., Schaefer M., Mueller-Her S.,
RA Gebel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Beger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Geoffau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gilbert F., Ayes S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rocher M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong I., Potashkin U.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin U.,
RA Sparovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 560-857 FROM N.A.
RC STRAIN=PR745;
RC MEDLINE=98162722; PubMed=9501991;
RA Yoshida K., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
RT CDNA."
RL DNA Res. 4:363-369(1997).
CC -1- SIMILARITY: Contains 2 SH3 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL162631; CAB83085.1; -
CC EMBL; D89164; BA13826.1; ALT_SEQ.
CC HSSP; P19174; 2SP.
CC GenDB SPombe; SPAPJ760.02C; -
CC InterPro; IPR002108; Actbind_cofin.
CC InterPro; IPR00108; Neu_cyt_fac2_2.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00241; Cofilin_ADF_1.
CC Pfam; PF00018; SH3; 2.
CC PRINTS; PR00499; P67PHOX.
CC PRINTS; PR00452; SH3DOMAIN.

```

```

DR ProDom; PD000066; SH3; 2.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
KW SH3 domain; Repeat.
FT DOMAIN 725 785 SH3 1.
FT DOMAIN 800 857 SH3 2.
SQ SEQUENCE 857 AA; 91137 MW; 247EFA78BAFE7097 CRC64;

Alignment Scores:
Pred. No.: 5,47e-05 Length: 857
Score: 214.00 Matches: 147
Percent Similarity: 33.90% Conservative: 72
Best Local Similarity: 22.76% Mismatches: 222
Query Match: 5.06% Indels: 207
DB: Gaps: 28

US-10-028-952a-4 (1-233) x APPL_SCHPO (1-857)
QY 11 CCCACTGTCCTTACCTACTATCCAGGAAACACAGCAAGGAGCGGCTTCCGGAAT 70
Db 295 ProthrrhrProasphrSphenylsLeuthrPro----- 305
QY 71 CAGCGGGAAGAAGAACCTGTGAGCTTGAGCCCGCCCGCCCGCCCGGCTGCCCTGA 130
Db 306 -----LeuthrlylsleuGluproGlnproProserGlyserProser 319
QY 131 AGGGGCGCGGCGGCGGCTCCCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCG 190
Db 320 LyslyserProvalsergluleuGlulLeuHsthrAlaGlysnValnsnleu-SerAl 339
QY 191 TCGT----- 194
Db 339 aArArAlaLeuPheglulyslysgluserSerThrlysnValGlnsnProvalse 359
QY 195 ----TTTTCACGTACCCCGGTAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 250
Db 359 rHisHleuLysSerProvalrArthrSerPheproProAlaserThrThrAlaserly 379
QY 251 CGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 310
Db 379 sglnAspserProserThrvalProvalaPlysglnGlulThrAlalysProleanty 399
QY 311 GCGAGTTGACTGGCGCGGTACCTGTCAACCGTACCGAGGTCTCCTAAGCGGAGCT 370
Db 399 sglnValserSerasnGlnThrSerAlaglIngluproArgInserValAlaAla 419
QY 371 CAGG-----GAGCAGGAACCTCCCGCGAGCAGAG 403
Db 419 uArgAlaArPheAlalysAlaValnsnValserGlnAsnAsnAspProthrPheProly 439
QY 404 GCGAAGACCTGCTGATCTTGATTTTCAGTACGAATACAGACCTGAAAGCGGGCTT- 462
Db 439 sThAlaAlalysle-SerSerPheAsnSerlysaAlaglYthSerPheAlalysProa 459
QY 443 -----CAGATCTCTGACCTTTGGTTTAAACAGAGAGTCA----- 504
Db 459 rGpProPheThrAsnAsnProsnProileSerAlaProGlnlyProThrSerGlyus 479
QY 505 -----GAAAAGTTACACAGGAGTAACTGCTT 532
Db 479 eTleuSerLeuAsnProProProAlaMetCProlysnValPheProGlnlyArgAspIleSer 499
QY 533 GTGGGGCGCCAGCGCTTACGACGACGCGCTTTGATCCTTCGATG-----TCGG 583
Db 499 eAlaSerGlnlyAlaAlaGlnProserValIlethrProserValProGlnProProa 519
QY 584 CTCTTCATCATATTGTGAGCAGATTC-----C 613
Db 519 lAlaProValValProGlnlAProserValHisGlnProProAlaAlaProValAlaP 539
QY 614 CAGCGTTGATGTTCACCCACTAATAGGAGAGGTAGCGGAGTTAGACCGTGTGAG 673

```


[illegible]

FT	DOMAIN	4634	4702	PRO-RICH .
FT	VARSPIC	1	305	Missing (in isoform 2) .
FT	VARSPIC			/FtId=VSP_008553.
FT	VARSPIC	306	672	PMELHPSMHSKCAVCYRACGAGSALINPNSWFENYSLIC HCHKAQGGQTITSAVQCHTPVCGRPSPPEGGDPTQEDPA LYVACQCGQKGVHTSMQPKPEGRLQCEARFLGACAYQLEP OLEAPLINEEMPLVPPEESPSPPEESPSPPEASRLSP PPEELPASPLPEALHTLSRLPEESPSPPEESPSPPESS PPEELSPSPPEESPSPSPALHTPLSPPEASPLSPPEES SPSPPEEPSPSPPEASRLSPPEESPSPPEESPSP PPASRLSPPEESPSPPEESPSPPEASRLSPPEESPLS PMASPPPEESPSPSPYSRLSPPYVYRSLSPPEESPSPS -> MSPPEESPSPPEASRLSPPEESPSPPEESPLSP PPPEASRLSPPEESPSPPEESPSPPEESPSPPEESPLSP SRLSPPEESPSPSPPEESPSPPEESPSPPEESPLSP PPPEESPSPPEESPSPPEESPSPPEESPSPPEESPSP HLSRLPEEPSPSPPEESPSPPEESPSPPEESPSPPEESPSP PEEPHLSPOAQEPHLSPOPELHLSPOTEEPHLSVPEEP LSPOPEHLSPOPEHLSPOPEHLSPOPEHLSPOPEHLS EKPEEPGCGCAPPEHLSPOPEHLSPOPEHLSPOPEHLS PPUSPLPEEPHLSPOPEHLSPOPEHLSPOPEHLSPOPEHLS A (in isoform 2) .
FT	VARSPIC	1454	1454	/FtId=VSP_008559.
FT	VARSPIC			E -> EGST (in isoform 3) .
FT	VARSPIC			/FtId=VSP_008560.
FT	VARIANT	4949	4949	R -> H (in dbSNP:3782356) .
FT	VARIANT			/FtId=VSP_017115.
FT	SEQUENCE	5262 AA;	564171 MM;	26B7C74CAD417E44 CRC64;

Alignment Scores:	
Pred. No.:	0.000648
Score:	195.50
Percent Similarity:	32.62%
Best Local Similarity:	23.89%
Query Matchn:	4.62%
DB:	1
US-10-028-952A-4 (1-3333) × MLI2_HUMAN (1-5262)	
	Length: 5262
	Matches: 178
	Conservative: 65
	Mismatches: 271
	Indels: 232
	Gaps: 33

```

QY      28  CTATTCAGCAAGAACCAACAGCCAGAGGAACGGGCTTGGCGGAATTAGG----- 75
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1851  lIeGIySerProthrlrPro-----AlaIyIeuSerThrSsrAlaSpGlyPheIeu 1866
      |||-----GGGAAGAAACACCTGTTTGAGCTTGAGCCCGCGGCG 111
QY      76  -----|||:::|||||
Db      1669  LysProProAlaGlySerValProGlyIProAspSerProGlyIuLeuPheLeuLysIeu 1888
      |||:::|||||
QY      112  CCCCCCGGATGTCCTCCCGGAGGGGCGCGGGGCGGATTCGCCGGCCCTGGCGGCGCGG 171
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1889  ProProGln-ValProAlaGlnAlaProSerGlnAspProheGlyLeuAlaProAla-- 1907
      |||::|||::|||::|||::|||::|||::|||::|||::|||
QY      172  TGAATACCACTACTCTCGATCGATTGTTTTCATGACATGACCCGGTAGGCGGGG-----GGCG 225
      |||||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1908  ---TyrProIeu-GluProAlaGlyPheProthrlAlaProProThrTyrProProTyPProS 1926
      |||::|||::|||::|||::|||::|||::|||::|||::|||
QY      226  AGCCCGGAGGAGCTCTCGCTTCTGAGCGCAAGCGCCCGCGCGCGCGCGCGGCGGCA 285
      |||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1926  erProthrlGlyAlaProAlaGlnProProMetLeuGlyAlaSerSerAlaProGlyAlaG 1946
      |||::|||::|||::|||::|||::|||::|||::|||::|||
QY      286  CCGCGTCGGGG-----ACAAGTCAGAGTGGGAGATTGACTGGGGCGGTACA 333
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1946  IyGln-ProIyGluPheHsThrThrProProGlyThrProAlaGlnAlaProSerThr 1966
      |||::|||::|||::|||::|||::|||::|||::|||::|||
QY      334  CCT-----GTCAACGGTAAACGACGTCTCTTAAGGCGAGCTACGGAGACAA 384
      |||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1966  ProAspProPheLeuLys-----ProAlaGlySerProSerLeuAspAspLeuAlaValPro 1983
      |||::|||::|||::|||::|||::|||::|||::|||::|||
QY      385  ACCTCCCGTGAAGCAAGAGGCAAAAGCTCGCTGATCTGATTGTTCA----- 432
      |||::|||::|||::|||::|||::|||::|||::|||::|||
Db      1984  GluSerProGlyValGlyIyGlyIyLeuAlaSerGluProLeuLeuSerProProProhe 2000
      |||::|||::|||::|||::|||::|||::|||::|||::|||
QY      493  -----GTACGAATTACAGACCGTGAAGAGCGGCGCTACAGATCTTCT 474
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db 2004 GlyUserArglysalaleuGluValylslysgluGluLeuGlyAlaSerSerProSer 2023
 QY 475 -----GACCTTTGGTTTAAAGCAGAGGTCAGAAAATTACCAAGAGATAA 525
 Db 2024 TyrGlyProProAlaLeuGlyPheValaSer-----ProSerSerGly 2038
 QY 526 CTGGCTGTGGCGCCAGCCTTCATAGCAGCGTCCTTTTTCATCCTTCATGTCGCT 585
 Db 2039 ThrHisLeuGlyGlyLeuGlyLeuLeuSerThrProAspValPheHisAlaProLeuThrPro 2058
 QY 586 CTTCCTATCATTTGAGAGAGAAATTCACCAAGCGTTGGATTGTTTCACCCATATATAGGA 645
 Db 2039 ArgAlaSerGlnValGluProGlnSerProGlyLeuGlyLeu----- 2072
 QY 646 ACGTAGCTGGGATTAGACCGTCGTAGAGACAGTTAGTTTACCTACTGATGATGTT 705
 Db 2073 -----ArgPro-GlnGluPro-----ProProAlaGlnAlaLe 2083
 QY 706 GTTGCCATGTATTCCTGCTCAGTACGAGAGAACCGAGGTTACAGACTTGGTGATG 765
 Db 2083 uAlaPro----- 2085
 QY 766 TGCTTGCTGAGAGCAATGGCGGAGACTACATCTGTGAGATTATGACTGAACGCT 825
 Db 2085 ----- 2085
 QY 826 CTAAGTCAGATCCCGCCAGCGGAGACGATACGAGACCGCCGCGGAGCTCGGTTGCC 885
 Db 2086 -----SerProProSerHisProAspIlePheArgProGlySerYrThrAspPr 2102
 QY 886 TCGATACCGGATCCCGCCGCTGTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 931
 Db 2102 o-TyrIaGlnProProLeuThrProArgProGlnProProProProGlnUserGly 2121
 QY 932 -----CCTCCACGCGCC-----CCGCGCGCGCGGAGGCGCGTCCCGCGCGCGCG 981
 Db 2121 salaleuProProArgSerLeuProSerAspProPheSerArgValProValaSerProGln 2141
 QY 982 GACCGCGGGTCGGGCGGAGAGCCCTCGTCCGCGGAAAGCGGCGCGCGGAGAGGC 1041
 Db 2141 nSerGlnSerSerSerGlnSerProLeuThrPro-----ArgProLeuSerAlaGluVal 2159
 QY 1042 GCGCGCCCTCGCCCGCTCAGCAGCAGCCGTCGTGGG-----GAACCTGGCGCTAAAACA 1098
 Db 2159 aPheCysProSerProValThr--ProArgPheGlnSerProAspProTyrSerArgPro 2178
 QY 1099 CCTCATCTCCAGT-----CCTCAGCCCTGCGC 1124
 Db 2179 ProSerArgProGlnSerArgAspProPheAlaProLeuHislyProProArgProGln 2198
 QY 1125 AAGCTGAGAGAGCCCTTCCTGAGAGAGAGCTCAACCCAGCAGACCCACTTGGAGAGA 1184
 Db 2199 ProProGluValAlaPheIleValaGlySerLeuIleHisThrSerLeuGlyAlaGlyGly 2218
 QY 1185 GAGCCAGCTGTGCTCATCTCAAGGCCAGGCGAGATTCCTCTGTGAGAGCCGCGCGCC 1244
 Db 2219 PheProAlaAlaLeuProAlaGlyProAlaGlyLeuHisAlaIleValaProSer-- 2237
 QY 1245 AGCACTCTCCATGCTGTGAGCAGAGAA----- 1274
 Db 2238 GlyGlnProProAsnPheValArgSerProGlyThrGlyAlaPheValaGlyThrProSer 2257
 QY 1275 -----GAGAGAGCTGTGATAG-----GAACCTCCAGAGAGAG 1307
 Db 2258 ProMetArgPheThrPheProGlnAlaValaGlyGluProSerLeuLysProProValPro 2277
 QY 1308 GAGACCTTCTACGAGCAGACCCCACTGGTGCAGAGCAGAGGTGTGCTGCAGACATTT 1367
 Db 2278 Gln-ProGlyLeuProProProHisGlyIleAsnSerHisPheGlyProGlyProThrIle 2297
 QY 1368 G-----ACCAACCAATTTCAGAGGCGCAGAGGCTC 1394
 Db 2297 uGlyLeuProGlnSerThrAsnYrThrValAlaThrGlyAsnPheHisProSerGlySe 2317

QY 1395 A--GTGGCAAGGGCTCT----- 1410
 Db 2317 rProLeuGlyProSerSerGlySerThrGlyLeuSerYrGlyLeuSerProLeuArgPr 2337
 QY 1411 -----GTGCCGTCCTCTGACACTACAGAGAGCGG-----AC 1445
 Db 2337 oProSerValaLeuProProProAlaProAspIleSerGluProTyrLeuSerHisGlyAl 2357
 QY 1446 GACACAGAGATCTCCCTTACCCCGAGAACCTCATCAGCGGATCGAGGTGATCGACGA 1505
 Db 2357 aSerGlnArgSerGlyIleThrSerProValaGluIleArg-----G1 2371
 QY 1506 GGCTGTGGCGGTGCTATGAGCGCGGATGCGCATTTGGCATGTCCCTGCCAATCAG-- 1563
 Db 2371 uAspProGlyThrGlyMetGly---SerSerLeuAlaThrAlaGluLeuProGlyThrG1 2390
 QY 1564 -----TGAGGCTCATGTAGTGAGGCTGAGGCGACATCTTGCCCTCTCAG 1613
 Db 2390 nAspProGlyMetSerGlyLeuSerGlnThrGluLeuGlyGlnArgGlnArgGlnArg 2410
 QY 1614 ACATGCTTCCTTAT----- 1629
 Db 2410 gLeuArgGluLeuLeuIleArgGlnGlnIleGlnArgAsnThrLeuArgGlnGluG1 2430
 QY 1630 -----GCTGGAAGAGAGCGCTGGAGATTGAC----- 1656
 Db 2430 uThrAlaAlaAlaAlaAlaGlyAlaValaGlyProProGlySerTrpGlyAlaGluProSe 2450
 QY 1657 -----ATTACACACTCTCCAGAAATAGACCC-----AGTAGAGATGA 1697
 Db 2450 rSerProAlaPheGluGlnLeuSerArgGlyGlnThrProPheAlaGlyThrGlnAspY 2470
 QY 1698 GGCCTCA-----GGCTCCCTCCCGCTTGGCAGACTAGCTGTACCCCAAAGCGAGC 1751
 Db 2470 sSerSerLeuValaGlyLeuProProSerLeuSerGlyProIleLeuGlyProGlySe 2490
 QY 1752 AATGCGCTGATTCACACACATCTCTCCTGATCCCGGAGCCCTCCAGAGACTTG 1810
 Db 2490 r-PheProSerAspAspArgLeuSerArgProProProProAlaThrProSerSerMet 2509
 Db
 RESULT 9
 ID 585 RAT STANDARD; PRT; 825 AA.
 AC C63003;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE 585 antigen.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Water; TISSUE=Brain;
 RA MEDLINE=96015159; PubMed=8557300;
 RA Suzuki E., Kojima N., Yoshimura K., Uemura K., Obata K., Akagawa K.;
 RT "Cloning and sequence analysis of cDNA for a possible DNA-binding
 protein 585 in the nervous system.";
 RL J. Biochem. 118:122-128(1995).
 CC - FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - TISSUE SPECIFICITY: Expressed in neurons.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D37934; BA07153.1; -
 DR PIR: J04163; J04163.
 KW DNA-binding; Nuclear protein; Antigen.
 SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD55BDF CRC64;

Alignment Scores:

Pred. No.:	Length:	825
Score:	193.50	Matches: 195
Percent Similarity:	30.39%	Conservative: 70
Best Local Similarity:	22.36%	Mismatches: 285
Query Match:	4.58%	Indels: 322
		Gaps: 43

US-10-028-952a-4 (1-2333) x SE5_RAT (1-825)

QY 2261 GGACACACCTGGGCGCTGCAGGCACTGCATGAG-----CCGGGTGGAGGACATGT 2211
 |||||
 Db 113 G1YThrlYlYrpgLYLeuSerProThrlAleuSerArgSpGlyYrGlu 132
 2210 TGTGCTCGCGTTTGTGAATACCCAGGCTGGAGCTTGCCATCTGCATCCCACTT--- 2154
 |||||
 Db 133 ThrGluThrProPhe-----SerProGluGlyAlaPheProGlyGlyGlyProAlaGlu 150
 2153 -----CCCATAGCCAGGACAGGACAGAGAAATGAGT----- 2118
 |||||
 Db 151 GluGluGlyValProArgProArgAlaProProGluProProArgProGlyAlaProArg 170
 2117 -----GGAGACACAGACAGAGGCTCCACACAGA 2091
 |||||
 Db 171 ProProArgProProGlyProLeuProLeuProLeuProGlySerGlnIuys---ProThrPhe 189
 2090 CAATTCCTCGCTGCCAAACCAACCATATCATCTCTGACTTGGTTCACAACTCTCTCTA 2031
 |||||
 Db 190 ValValGlnValSerThrGlnGlnLeuLeuSerThrGlyGly-----CysAsp 206
 2030 AAACATCTCTCTAGCTCTACTGTCTCCAGAGGT-----CAT 1995
 |||||
 Db 207 LysGlu-----ProProArgGlyGlnGlnValAspThrArgGlyAspArg 221
 1994 TCTTAACAGTGTGGCCACCACTCCCTTAAAGCAAGAAAGAAAGAAAGCAATATA 1935
 |||||
 Db 222 ThrGlnGlu-GlyGlyGlnIuysProArgGlnGlnArgGlyGlyProArgProGlnGln 241
 1934 TGCAGAGGACGACGACGAGCTCCAGATGCCCTGGCTCAGAGCCCTCCCGGACGCTGG 1875
 |||||
 Db 241 YProAspIleProGlyGlnGlnGlnIuysProGlnGlnIuysProSerSerGlyArg 261
 1874 CAGT-----GGCACTCAAGGCTTGGCCACAGCA 1845
 261 YAspSerValGlyGluArgGlnAlaArgSerProGlnIuysGlnGly-----G1 277
 1844 GGGCTTGGCTCAGATCTCT-----GTCAAGGACAGAGCCCAAGCTCTG--- 1799
 |||||
 Db 277 uGlyGlyGlyGlnIuysProGlyIleSerGlyGlnIuysArgGlnIuysProGlyGlnIuys 297
 1798 -----GAGGATCGGGGATGATGCTGCAGAGGATGTGTGGATCAACAGGCAATT 1752
 |||||
 Db 297 YAlaAspValProArgGlyArgGlyGlnGlyAlaGlyGlnIuysArgValPro-- 316
 1751 GCTGCATTTGGGCTGACAGGCTGATGCTGCAGAGCCGAGAGGACCTTGAGGCTCATCC 1692
 |||||
 Db 317 -----LysAspArgGlyGlnGlyArgGlnIuysProGlnIuysAlaGlnG1 333
 1691 TCAC-----TGGGGCTCTATTCTGGAAGAGTGTCAATGTCATG 1650
 |||||
 Db 333 uHisGlyGlnAlaAlaArgAspIlePThrSerGlnIuys--ProArgThrLeuGlyGlnAsp 353
 1649 CCCAGGCTCTCTCTTCAGCAATAAGAGACCATGTCTGAGAGGGAAGGCAAGATGTC 1590
 |||||
 Db 353 AlaAspArgGlySerSerSerArgAspAla-----AlaGlySerSerProGly 370
 1589 CCCTC-----AGCTCACTCA----- 1573

Db 370 lAleuArgGlySerLeuAlaProGlnArgLeuGlyAspGlyProTrpProAlaTrpPro 390
 QY 1572 -----TAGCTCCACAGTATGGACAG 1551
 Db 390 eProGlnGlnArgGlnProGlyProArgAspArgValGlnSerProArgGlnIuysGly 410
 1550 GAACA-----TCCCAAAATGCGCATCCGACCAT 1521
 |||||
 Db 410 LyrThrGlnSerProArgGlyYrpgLulAlaGlyProArgGlnIuysProSerProGlyG 430
 1520 GCCAGGCCACAGCTTGTGTGATCATCTGCATCCCGTATGAGTTCCTGGGGTCAAA 1461
 |||||
 Db 430 lYArgGlyAspGlyProArgArgArgProArg-----LysArgArgGlyArgL 446
 1460 GAGATCTCTGTGTGCTGCGCTGCTGTAGTGTGTACAGAGGACAGGACAGCCCTTG 1401
 |||||
 Db 446 YGlyArgMetGlyArgGlnLeuGluThrThrAlaThrSer-----AlaSerA 462
 1400 CCCACTGAGCCCTTGCGCTGATGTGTGTGATGTCTCAATGTCTCAGAGCCACACCTTCTG 1341
 |||||
 Db 462 lAthrGlyGlyProAlaGlnGlnAlaGlyAlaSerAlaProGlnGlyGln---AlaGlyG 481
 1340 CTCGACACAGTGGGGCTGCTGTGTAAGGTCTCTGCTGTGAGGTTCTCATACAGAGC 1281
 |||||
 Db 481 lYgIYProArgGlyArgAlaArg----- 488
 1280 CTCTCTTCTGCTGACACACATGAGAGAGTGTGGGCGCGGCTCTTCACAGAGAG 1221
 |||||
 Db 489 -----GlyProArgGlnGlnAlaArgArgArgGlnGlyProGln-Arg--- 502
 1220 ATCTGCGCTGGGCTGTAGATGACAGAGCTGGCTCTCTGCAAGTGGGCTGTGTTG 1161
 |||||
 Db 502 ----- 502
 1160 GGTGAGCTGCTTCTGACAGAAAGGCTCTGAGCTTGCAGAGCTGAGATGAGAT-- 1104
 |||||
 Db 503 -----ArgArgGlyProProGlnAlaGlyGlnGlnGlyProGlnIuysAla 517
 1103 -----GAGAGTGTGTAGCGGACAGTCTCCACAGAGTGGGGCTGAGAGCG 1056
 |||||
 Db 518 ThrLeuValLeuGlyLeuGlyThrThrSerGlyGlnIuysArg-----AlaAspGln 534
 1055 GCGAGGGGCGCGCGCTTCCGCGCGCGCGCGCTTCCAGAGAGAGGCACTCGCA 996
 |||||
 Db 535 SerGlnThrLeuProAlaLeuAlaGlyAlaProThrAla-----HisAlaHisAlaVal 552
 995 CCGAGCCCGGCTCCCGGCGCGCG-----CGGGCAGCGCGCTCCCGCGCGC 948
 |||||
 Db 553 ProGlyProGlyProAlaAlaAlaThrLeuGlyArgGly-Arg----- 567
 947 GCGGGGCGCGTGAAGGGGGGCGCGCGCGCGCGCGG3-----ACAGGCGGGAGC 897
 |||||
 Db 568 ArgGlySerIleThrArgGlyGlyArgArgGlyGly-GlyAlaGlyAlaSerGlyGlyGly 587
 896 CGGCTATCCAGAGCCACCGAG-----CTCCGGGGGCTCGGATTCGTCGCGTGG 843
 |||||
 Db 587 rGgIyGlyArgGlyArgGlyArgGlyArgGlyArgGlySerGlyLeuSer----- 603
 842 GCGGATTCGTAGAGGCGTTCAGTCATATCCACAGATGATGATTCGCCCCATT 783
 |||||
 Db 604 -----G 604
 782 GGGCTCTACCCACAGCATACCAATGTCTGACCTGGCTCTCTGCTACTAGC 723
 |||||
 Db 604 lYThrArgGlnAspAlaGlySerProser-----AlaArgArgGlyGlyG 619
 722 AGGATTACATGGCAACAACATCATCATGAGGTAAACTTAACCTGTCTCAGACAGCT 663
 |||||
 Db 619 lNArgArgArgGlyHisGlyPro-----ProAlaAlaGlyAla 632
 662 CTAAATCCAGCTCAGCTTCCCTATTAGTGGGTGAACAATCCAGCTTGTGGAATTCTGC 603
 |||||
 Db 632 lAGlnValSerThrArgGly----- 638

[illegible]

```

RESULT 10
YHLL_EBV
ID_YHLL_EBV      STANDARD;      PRT;      660 AA.
AC_P03181.
DT_21-JUL-1986 (Rel. 01, Created)
DT_21-JUL-1986 (Rel. 01, Last sequence update)
DT_15-JUL-1998 (Rel. 36, Last annotation update)
DE_Hypothetical BHLPI protein.
OS_Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC_Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC_Gammaperovirinae; Lymphocryptovirus.
CX_NCBI_TaxID=10377;
RN_
RP_SEQUENCE FROM N.A.
RX_MEDLINE=84270667; Pubmed=6087149;
RA_Baer R., Bankier A.T., Birgin M.D., Deininger P.L., Farrell P.J.,
RA_Gibson T.J., Hatfull G., Hudson G.S., Satchell S.C., Segun C.,
RA_Tufnell P.S., Tufnell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL_Nature 310:207-211 (1984).
-----
CC_CC The SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).

```

CC	EMBL: V01555; -; NOT_ANNOTATED_CDS.
DR	PIR: A03742; OOE83.
KM	Hypothetical protein; Early procein; Repeat.
FT	DOMAIN 149 646 4 X 125 AA TANDEN REPEATS
FT	REPEAT 149 273 1.
FT	REPEAT 274 398 2.
FT	REPEAT 399 523 3.
FT	REPEAT 524 648 4.
SQ	SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Alignment Scores:

Pred. No.:	0.000985	Length:	660
Score:	191.50	Matches:	169
Percent Similarity:	27.86%	Conservative:	38
Best Local Similarity:	22.75%	Mismatches:	213
Query Match:	4.53%	Indels:	324
DB:	1	Gaps:	34

US-10-028-952A-4 (1-2333) X YHL1_EBV (1-660)

QY 2 CCACTATACAGCAAAACACACAGCAAGGAGCGGCTGGCGGAATACAGCGAGGAAAGA 83
 Db 95 ProAlaGlueHAlaAaphAlaAlaIserAsnProthrgLyGlyCySserAspProGln 114
 QY 84 AGACCTGTGAGCTTGGGCCCCCGGCGCCCCCGGTGTCGCCGA----- 131
 Db 115 Arg-----SerProAlaGthrArgGlnAla 122
 QY 132 -----GGGGCGCGGGGCGGGGCGCGGCC 158
 Db 123 GLYTYrAlaLeuGlyGlyLeuIyserylserAlaLeuIyserylserArgLy-----ProArgPro 140
 QY 159 TGGGGCGCGCGGATGAAATACCACTACTGTATCGTTTTCTACTGACCG----- 209
 Db 141 ---HisProAlaAphGlnValGlnIlePheSerAlaArgAsnProGlyCySPromArgThrTyr 159
 QY 210 GTAGAGCGGGGGGGGCGAGCCCGGAGGGGCTCTGGCTTGTGGCGCGAAGCG----- 260
 Db 160 ArgArgArgSerLeIyAlaGlnArgGlyIleSPromProProGlyAlaGlyGlnArgProSer 179
 QY 261 ---CGGGCGCGCGCGCGCGCGCGCGCGACCC---GCTCCGGGAGCAAGTCCAGGTGGGGA 314
 Db 180 GlyProthrgLyGlyArgProAlaAla-ProGlyAlaAProGlyThr---ProAlaAlaArg 198
 QY 315 GTTTGACTGGGCGCGGTACACCTGTGTAAAGGTAAAGCAGGTGTCTTAAGCGAGCTCAGG 374
 Db 198 cGlyProGlyGlyGlyAlaAlaVal-----ProSerG 209
 QY 375 GAGGACA---GAAACCTCCCGTGGAGCAGAAAGGCAAAAGCTCGTTGATCTGATTTTC 431
 Db 209 yAlaThrProHisProGlyArgLy----- 217
 QY 432 AGTACGAATACAGACCGGTAAAGCGGGGCGTCAGATCTCTTGACCTTTGGTTTAA 491
 Db 218 -----SerGlyProAlaSPromProAlaAlaAlaArgLeuP 230
 QY 492 GCAGAGGTGTAGAA---AACTTACCAAGAGATTAAGCTGTGTGGCGCCAGGCTT 548
 Db 230 oProGlyArgGlnGlnIleProArgLeuProGln----- 240
 QY 549 CATAGGAGCGTCGCTTTTGATCTTCGATGTGCGGCTCTTCTCATCTGTGAAGCAGAA 608
 Db 241 -----AspIleuAlaAlaGlnIle 247
 QY 609 TTCACCAAGCGTTGATTTGTTCACCCACTAATAGGGAAGTGAAGTGGATTAAACCGTC 668
 Db 247 gCySProm-----AlaGlyProPProth 255
 QY 669 GTAGAGACAGTTAGTTTACCTACTGATGATGTGTGTTCAGCANGTAATCCGCTCAG 728
 Db 255 T----- 255

```

QY 729 TACGAGAGAACCGGAGGTTCAACATTGTTGTTGCTGCTGAGAGCCAAATGG 788
DB 255 -----
QY 789 GCGAAGCTACATCTGTGGATTATGATGAACGCTTAAGTCAGAAATCCGCCAAGC 848
DB 255 -----
QY 849 GGAACGATACGCGAGCGCGGAG-----CCTGGTTGGCTCGGA 890
DB 256 ---ArgSerGlyAlaAlaAlaGlnArgThrHisArgProProGlyCysProArgSe 274
QY 891 TAGCCGCTCCCGCGCTGTCGCGC----- 915
DB 274 TAlaArgAsnProGlyCysProArgThrTPAArgArgSerGlyAlaGlnArgGlyHi 294
QY 916 -----CGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 953
DB 294 sProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaPr 314
QY 954 GCGAGGCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTC 1013
DB 314 ogGlyAlaProGlyThProAlaAlaProGlyProGlyGlyAlaAlaValProSerG 334
QY 1014 CTGG-----GAAACGGGCGCGCGCGCGGAAAGCGCGCGCGCGCGCGCTCA 1061
DB 334 yAlaThrProAlaProGlyAlaGlySerGly---ProAlaAspProProAlaAlaAlaAr 353
QY 1062 CGCAGCGCGACGTTGCGGGGAACCT----- 1086
DB 353 gleuProProGlyAlaArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCy 373
QY 1087 -GGCGCTAAACACACCTTCATCTCAGTCTCAGCC-----TGGCAAGCTG 1130
DB 373 sProAlaGlyProProProThr-ArgSerGlyAlaAlaAlaGlnArgThrHisArgArgP 393
QY 1131 AGAGAGCCCTCTCTGCGAAGACGCTCACCCCAACGAGAACCCACTTGGCAGAGACCA 1190
DB 393 roProGlyCysProArgSerAlaArgAsnProGlyCysProArgThrTPAArgArgArgS 413
QY 1191 GCGTGTGCGCATCTCAAGAGCCGAGGAGATCTCCTGAGAGCGCGCGCGCACT 1250
DB 413 et-----GlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProS 429
QY 1251 CCTCCATGTTGTGCA-----G 1268
DB 429 ergGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThProAlaAlaProG 449
QY 1269 GCGAAGAGGAGGCTGTATGAGAACTCCAGAGAGAGAGACCTTACGAGCAGCC 1328
DB 449 lProGlyGlyGlyAla-----AlValP 457
QY 1329 CCACTGTGAGCAGCAGAGGTGCTGCTGAGCAATTGACCAACACAT----- 1378
DB 457 roserGlyAlaThr-----ProHisProGluArg 467
QY 1379 --TACGAGCGAGGCGCTCAGTGGCGAGGCTGTGTCGCG----- 1417
DB 467 lySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProA 487
QY 1418 -----TGCCTGTA----- 1426
DB 487 rgleuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgS 507
QY 1427 -----CGACTACAGCAGAG-----CGAGCAGCAGAGAT 1456
DB 507 ergGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArg 527
QY 1457 CTCCTTTGACCCGAGAACCTCATCAGCGGCGATG----- 1491
DB 527 snProGlyCysProArgThrTPAArgArgSerGlyAlaGlnArgGlyHisProProP 547
QY 1492 --AGGTATGACAGAAAGGCTGTGCGCTAGTGGCGCGATGGCCATTTTGGCATGTT 1549

```

```

DB 547 roGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla-----AlaProG 565
QY 1550 CCTGCCCACTACGTGAGAGTCAATTGATGAGGCTAGGGCACATCTTCCCTCC 1609
DB 565 lAlaProGlyThProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyA 585
QY 1610 TCAGACATGGCTCTTATTGTGGAAGAGAGGCTGGGAGTTGACATTCACACTT 1669
DB 585 la-----ThProHisProG 590
QY 1670 CAGAAATAGACCCCGATGAGATGAGGCTCAGG-----CTCCCTCCGGCTTGCA 1726
DB 590 lArgGlySerGlyProAlaAspProProAlaAlaArgLeuProProGluArgGlnG 610
QY 1727 CTAGCGCTGACCCCAATGAGCAGCAATGCGCTGTGATTCACACATCTTCTGCAT 1786
DB 610 lProAlaGlyLeuProGlnAspLeuAlaAla-----GlnArgCysProAlaG 626
QY 1787 CCCCCGACCTCCGAGACAGCTTGTGCTTGTGCGCTGACAGATTAAGCCAGCCCTG 1846
DB 626 lPro---ProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProG 645
QY 1847 CTGCT 1851
DB 645 lCys 646

RESULT 11
ICP0_HSV11 STANDARD; PRT; 775 AA.
AC P08353;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trans-acting transcriptional protein ICP0 (Immediate-early protein IE110) (VNM110) (Alpha-0 protein).
ID ICP0 OR IE110.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=283594;
RA McGeoch D.J., Delymple M.A., Davison A.J., Dolan A., Frame M.C.,
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87059760; PubMed=3023529;
RA Perry L.J., Rixon F.J., Everett R.D., Frame M.C., McGeoch D.J.,
RT "Characterization of the IE110 gene of herpes simplex virus type 1."
RL J. Gen. Virol. 67:2365-2380 (1986).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89036163; PubMed=2846760;
RA Perry L.J., McGeoch D.J.,
RT "The DNA sequences of the long repeat region and adjoining parts of
RT the long unique region in the genome of herpes simplex virus type
RT 1."
RL J. Gen. Virol. 69:2831-2846 (1988).
CC -!- FUNCTION: ICP0 IS A GENERAL TRANSACTIVATOR OF ALL THREE CLASSES OF
CC HSV GENOMES WHICH FUNCTIONS SYNERGISTICALLY WITH ICP4. OFTEN TERMED
CC AS A PROMISCUOUS TRANSACTIVATOR. MAY PLAY A ROLE IN THE BALANCE
CC BETWEEN THE LATENT AND LYtic STATES, REACTIVATING LATENT HSV. ICP0
CC MIGRATES TO THE ND10 NUCLEAR BODIES IN THE EARLY TIMES OF
CC INFECTION AND DISRUPTS THEM. BINDS TO HAUSP AND MAY MODIFY ITS
CC DEGRADITINATING SUBSTRATE SPECIFICITY OR ACTIVITY ON VIRAL OR
CC CELLULAR TARGETS, LEADING TO AN INCREASED OR DECREASED STABILITY
CC OF THESE PROTEINS.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.

```


Dd		583	aSerSerSerAlaSerAlaSerSerGlyGlyAlaValGlyAlaLeuLysIleGlyArg-----	601
Oy		1653	TGAACATTTCAGCACTCTTCCAGGAATAGAGCCCCCAGTAGAGATAGGCTCACGGGTCCC	1711
Dd		602	-----GNGJUGLUthr-SerLeuGLYProArgAlaAlaSerGlyProArgGlyProA	619
Oy		1713	TCCGGCTTGCGACACTCAGCCTGTGCACCCCAATGCA--CAATGCCCTGTGATTTCC	1765
Dd		619	rGlysCysAlaArgLyThrArgHLeaIaGluThrSerGlyAlaValProIaGlyIYL	639
Oy		1770	ACGACATCTTCCTTCGACATCCCCGACCCCTCCAGAGAGCTTGAGCTTGGCTCCGCC	1822
Dd		639	euthrArgLyTrileuProlIeSerGlyAlaSerSerValAlaIaLeuSerProTyrrValA	659
Oy		1823	--GACAGATATCTAGCCCAAGCCCTGCCTGTg	1852
Dd		659	snLystrHrIleThrGlyAspCysLeuProIle	669
RESULT 12				
ID	CALL_CHICK	STANDARD;	PRT,	1453 AA.
AC	P02457;			
DT	21-JUL-1986	(Rel. 01, Created)		
DT	01-OCT-1989	(Rel. 12, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Collagen alpha 1(I) chain precursor.			
GN	COL1A1			
OC	Gallus gallus (Chicken).			
OC	Eukaryote; Metacoela; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
CC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE OF 1-153 FROM N.A.			
RX	MEDLINE=88056316; PubMed=36788934;			
RA	Fisher M.H., Boedtker H., Doty P.;			
RT	"Construction and characterization of cDNA clones encoding the 5' end			
RL	of the chicken pro alpha 1(I) collagen mRNA.";			
RL	Gene 56:71-78(1987).			
RN	[2]			
RP	SEQUENCE OF 1-144 FROM N.A.			
RX	MEDLINE=88007542; PubMed=2820966;			
RA	Fisher M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;			
RT	"Unusual DNA sequences located within the promoter region and the			
RL	first intron of the chicken pro-alpha 1(I) collagen gene.";			
RL	J. Biol. Chem. 262:13323-13332(1987).			
RN	[3]			
RP	SEQUENCE OF 152-1187.			
RX	MEDLINE=82231995; PubMed=7093229;			
RA	Hinberger G.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,			
RT	Kang A.H., Gross J.;			
RL	"Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the			
RT	complete primary structure of the helical portion of the chick skin			
RL	collagen alpha 1(I) chain.";			
RL	Biochemistry 21:2048-2055(1982).			
RN	[4]			
RP	SEQUENCE OF 1200-1205.			
RX	MEDLINE=72243016; PubMed=5047697;			
RA	Eyre D.R., Glimcher M.J.;			
RT	"Evidence for a previously undetected sequence at the carboxy-terminus			
RL	of the alpha 1 chain of chicken bone collagen.";			
RL	Biochem. Biophys. Res. Commun. 48:720-726(1972).			
RN	[5]			
RP	SEQUENCE OF 981-1453 FROM N.A.			
RX	MEDLINE=81160715; PubMed=6927845;			
RA	Fuller F., Boedtker H.;			
RT	"Sequence determination and analysis of the 3' region of chicken pro-			
RT	alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids			
RL	including the carboxy-terminal propeptide sequences.";			
RL	Biochemistry 20:996-1006(1981).			
RN	[6]			
RP	SEQUENCE OF 1311-1453 FROM N.A.			

```

CC MEDLINE=80134546; PubMed=6987088;
RA Showalter A.M., Pasciotta D.M., Eikenberry E.F., Yamamoto T.,
RA Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen CDNA-fragment coding for the
RT carboxyl end of pro alpha 1(I)-chains."
RL F855 Lett. 111:61-65 (1980).
CC -1- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -1- PFM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains. Pro-1153 is the only 3-hydroxypro and the only
CC hydroxylated proline in position X.
CC -1- SIMILARITY: Contains 1 WPMC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce).
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL: M17839; AAA48704.1; -.
DR EMBL: M17838; AAA48704.1; JOINED.
DR EMBL: Y00401; CAA23695.1; -.
DR EMBL: M10571; AAA48671.1; ALT_SEQ.
DR EMBL: M17607; AAA48672.1; -.
DR PIR: A27179; A27179.
DR PIR: I50629; I50629.
DR InterPro: IPR008161; C1g_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR002181; Fibtrogen_C.
DR InterPro: IPR001007; WPMC.
DR Pfam: PF01410; COLF1; 1.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF00093; WMC; 1.
DR ProDom: PD000007; C1g_helix; 2.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SMO0038; COLF1; 1.
DR SMART: SMO0214; WMC; 1.
DR PROSITE: PS01208; WMC_1; 1.
DR PROSITE: PS0184; WMC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
FT DONAIN 31 89 WPMC.
FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
FT MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1153 1153 HYDROXYLATION.
FT CONFLICT 1187 1187 F -> I (IN REF. 5).
FT CONFLICT 1441 1441 Q -> H (IN REF. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Alignment Scores:
Pred. No.: 0.00141 Length: 1453
Score: 189.00 Matches: 184
Percent Similarity: 30.49% Conservative: 34
Best Local Similarity: 25.73% Mismatches: 236
Query Match: 4.47% Indels: 262
Gaps: 42

```



```
QY      142 CCCCCGCCCCCTGCGGGGACACCGGGGGGGCC-----108
      : : ||||| ||| ||| ||| : :
Db      1300 erglythrProSerArglyGlyArgSerGlySerAentrpGlyArgGlySerAsnMetA 1320
QY      107 --GGGGGCCCAAGCTCAACAGGGTCTCT 81
      : : ||||| ||| ||| : :
Db      1320 snserGlyProProArgArgGlyAlaSer 1329
```

Search completed: April 1, 2004, 15:54:43
Job time : 77 secs


```

QY 961 CGCGTGGCCCGCGCGCGCGAGACCGAGGCGCGGTGCGAGTGCCTTCGCTCGGAAA 1020
Db 371 -----ProAlaProAlaProThrProSerAlaGlyLeu---ValGlySerAlaGlySGLu 387
QY 1021 ACGGGGCGCGCGCGGAAAAGCGCGCGCGCGCTCGCGCTCAAGCAGCAGCGCTTCGTGGG 1080
Db 388 LeuPheSerSerLysProSerGlyProValLeuProLys-----ProGlnValAlaSerGly 405
QY 1081 GAACCTGGCGCTTAACACACCTCCCTCCACCTCTTCAGCTTCGAGCACTGAGAGCCCT 1140
Db 406 SerProLysLysTrpProPro--ValGlyThrThrSerProPro-----ArgGluProV 423
QY 1141 TCCTGCAGAACAGCACTCAACCCAGCAGAG-----ACCCACTTTG 1179
Db 423 al-----AsnArgLeuThrGluProAlaProAlaProSerThrTyThrProLysProIleA 441
QY 1180 GCAGAGAGCGCGCTGCTCCCTCAATCCAGCGCCAGGAGATCTCCCTGCT----- 1229
Db 441 laTyGluProGluProMetValTyTrpLysProGluAlaMetLysProAlaValSerTyra 461
QY 1220 -----GAGGAGCGCGCGCGCGCGACT----- 1250
Db 461 spAlaTyGluGluProProAlaGluProAlaProProThrPheLeuAlaProThrProV 481
QY 1251 -----CCTCATGCTCTGTCGTCAGAGCAGAAAGAGAGAGCTGTCTATG 1290
Db 481 alIleAlaProProProProGluProGluProThrProAlaProSerHisTyraLaserGlnTyra 501
QY 1291 AGGAACCTCTCCAGAGAGAGAGAGCTTC----- 1316
Db 501 spAlaProProValHisGluSerPheGluProAlaValProProValSerAlaProSerH 521
QY 1317 -----TAGAGCAGACCCCGCAGCTGTCGTCAGCAGCAGAGCTGCTGAGC 1362
Db 521 lTyraLaserGlnTyraSpAlaProProGluProIleAspSerHisSerSerSerG 541
QY 1363 ACATTGACACACACATTCAGGCGCGAG----- 1388
Db 541 lInleuProAlaHisIleAlaSerGlnTyraAspMetProProValMetProGluProV 561
QY 1388 ----- 1388
Db 561 alPheAlaProLysSerSerProIleLysValAlaAlaProProIleAspGlnTyraSP 581
QY 1389 -----GGGCTCACTGGGCGAAGGCGCTGTGCGCGCTGACCTGACACATCAACAGGCGAG 1440
Db 581 heProProAlaValAlaGluGlnAsn-----AlaMetAlaLeuThrAspTyraGlnAlaA 599
QY 1441 CCGACGACACAGAGATCTCTTTGACCCCGAGAACCTCATCAACGCGCATCGAGCTGATCG 1500
Db 599 laAspAspThrGluLysSerPheAspProAspAspIleIleThrAspIleAspGlnValA 619
QY 1501 AGGAAGCTGCTGCGCTGCTATGGCGCGAGAGCGCATTTTGAGTGTCCCTGCACT 1560
Db 619 spSerGlyTyThrTyraGlyAlaProSerGlyAlaGlyValGlyLeuPheProAlaAsnT 639
QY 1561 ACGTGAGCTCATT 1574
Db 639 yTValLysLeuIle 643

```

```

A:Cross-references: EMBL:U16663; NID:932054; PID:CA934651.1; PID:932055
R:Yamanashi, Y.; Okada, M.; Semba, T.; Yanori, T.; Umemori, H.; Tsunashima, S.; Toyoshima
Proc. Natl. Acad. Sci. U.S.A. 90, 3631-3635, 1993
A:Title: Identification of H81 protein as a major substrate of protein-tyrosine kinase (
A:Reference number: A47478; MUID:93234551; PMID:7682714
A:Accession: A47478
A:Status: preliminary
A:Molecule type: protein
A:Residues: 4-19, 'XXX', 23-26, 'X', 79-93, 'X', 95, 'X', 134-146, 208-223, 274-284, 'X', 286, 'X', 2
A:Experimental source: Daudi, B-lymphoblastoid cells
A:Note: sequence modified after extraction from NCBI backbone
C:Superfamily: SH3 homology
F:435-482/Domain: SH3 homology <SH3>

Alignment Scores:
Pred. No.: 4,77e-11 Length: 486
Score: 277.00 Matches: 87
Percent Similarity: 42.97% Conservative: 23
Best Local Similarity: 33.98% Mismatches: 85
Query Match: 6.55% Indels: 61
DB: 2 Gaps: 10

US-10-028-952A-4 (1-2333) x 507633 (1-486)
QY 938 CGCGCCCGCGCGCGCGGAGGCGCGCTGCGCGCGCGCGCG-----GACCGAGGT 991
Db 250 ArgLysArgGluGluGluGluGluValAlaGlnValAlaArgGluGluGluLys 269
QY 992 CCGGTG-----CGAGTCCCTTGTCTGCGGAAAAGCGCGCGCGGCGCG 1045
Db 270 AlAlaValThrLysArgSerProGluAlaProGluProValIleAlaMetGluGluProAla 289
QY 1046 GCGCCCTGCGCGCGCTCAACGACGACCGAGCTGCGGGAACCTGCGCTAAACCACTCCAT 1105
Db 290 ValProAlaProLeuProLys----- 297
QY 1106 CTCGACTCTCAAGCTGCGAGCTGAGAGAGCCCTTCTGCGAAGACGCTACCCAGC 1165
Db 298 lIleSerSerGluAlaThrProProValGlyThrProProSerSerGluSerGluProVal 317
QY 1166 AGAGCCCACTTGGCAGAGAGAGAGCTGCTCCATCTCAAGGCGCGAGGCGAGTTCGC 1225
Db 318 Arg-Thr-----SerArgGluHisProValProLeuLeuProIleArgGlnThrLeuPr 335
QY 1226 TGCT-----GAGAGCCGCGCGCGCGAGCACTCTCCATCTG-----GTGAGGC 1270
Db 335 ogLysAspGluGluPro--ProAlaLeuProProArgThrLeuGluGluGlnValA 354
QY 1271 AGAAGAGAGGCTGTATGAGGAACCTCCAGAGCAGAGACCTTCTACAGACAGCCCGC 1330
Db 354 lGluGluGluProValTyGluAlaGluProGluProGluPro-----GluProGluPr 372
QY 1331 ACTGTGACAGCAAGGTCTGCTGAGCAATGACCAACCATTCAG----- 1382
Db 372 ogLysGluAspAspTyraGluAspValGluGluMetAspArgHisGluGluGlnAspG 392
QY 1382 ----- 1382
Db 392 uProGluGlyAspTyraGluGluValLeuGluProGluLysSerSerPheSerSerAla 412
QY 1383 -----GGCCAGGGGCTCAAGTGGCGAAGGCTCTG 1411
Db 412 uAlaGlySerSerGlyCysProAlaGlyAlaGlyAlaGlyAlaValAlaLeuGlyIle 432
QY 1412 TGCCCGTCCCTGTAGACTACAGCAGCAGCGAGCAGCAGACAGAGATCTCTGACCCCGA 1471
Db 432 lAlaValAlaLeuTyraAspTyraGlnGluGlySerAspGluLeuSerPheAspProAs 452
QY 1472 GAACCTATCAAGGAGCTGAGTTCAGAGAGGCTGCTGCGCTGAGTGGCGCGGA 1531
Db 452 pAspValIleThrAspIleGluMetValAspGluGlyTyThrTyraGly---ArgCysH 471
QY 1532 TGCCATTTTGGCATGTCCTGCGCAACGACGAGCTCATTCAG 1577

```



```
Db      471 sGlyHisPheGlyLeuPhe-ProAlaAsnTyrValIysleuLeuGlu 486
```

RESULT 3

LckBp1 protein - mouse
N:Alternate names: Hs1
C:Species: Mus musculus (house mouse)
C:Date: 29-May-1998 #sequence revision 29-May-1998 #text_change 05-Nov-1999
C:Accession: I49760; S57244; S52474
R:Kitamura, D., Kaneko, H., Tanituchi, I., Akegi, K., Yamamura, K., Watanabe, T., Blochman, B.ophy's. Res. Commun. 208, 1137-1146, 1995
A:Title: Molecular cloning and characterization of mouse Hs1.
A:Reference number: I49760; PMID:1553527
A:Accession: I49760
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-486 <RES>
A:Cross-references: GB:D42120; NID:g806521; PIDN:BA07701.1; PID:g806522
R:Takemoto, Y., Furuta, M., Li, X.K., Strong-Sparks, W.J., Hashimoto, Y.
EMBO J. 14, 3403-3414, 1995
A:Title: LckBp1, a p10line-rich protein expressed in hematopoietic lineage cells, directed
A:Reference number: S57244; MUID:95354658; PMID:7628441
A:Accession: S57244
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-486 <TAK>
A:Cross-references: EMBL:X84797; NID:g683480; PINN:CA59265.1; PID:g683481
C:Superfamily: SH3 homology
A:36-463/Domain: SH3 homology <SH3>

QY	1233	GAGCGGGGCGCCAGACACTCTCCTCATGTCTGTGTGCAGCAGAAGGAGGCGTGTATGATG	1292
Db	361	GlutProValylArgIuaIaaIaIaProGluLeu--GluProGluProGluProAspTyrGlu	379
QY	1293	GAACCTTCAGACGACGAGACCTTCTACGACAGC-----	1322
Db	380	ProGluProGluThrGluProAspTyrGluAspValGlyIuLeuAspArgGlnAspGlu	399
QY	1326	-----CCCCAGCTGGTGAC	1344
Db	400	AspAlaGluGlyAspTyrGluAspValLeuGluProGluAspThrProSerLeuSeryr	419
QY	1341	CAGCAAGATGGTGGCTCTGACGACATTGACCACACATTTCAGGGCCAGGGGCTCAGTGGG	1400
Db	420	GlnIleAlaIyProSerAla-----GlyAlaGlyGly	429
QY	1401	CAAAGGCTCTGTGGCCCGTGCCTCTGTACGACTACGAGGACGCCAGCACACAGAATCTCC	1460
Db	430	AlaIylIleSerAlaIaIeAlaIeAlaLeuTyrAspTyrGlnGlyGlySerAspGluLeuSer	449
QY	1461	TTTAACCCCAGAAACCTTCATCAACGGGATCGAGGTATCGAAGGATGGTGGCGTATGC	1522
Db	450	PheAspProAspAspIleIleThrAspIleGluMetValAspGluGlyTyrTrpArgGly	469
QY	1521	TATGGGCGGATGAGCTTTTGGAGATCTTCCTCCACAATCGATGGAGCTATT	1574
Db	470	Gln--CysArgGlyAlaPheGlyLeuPheProAlaAsnTyrValIylsLeuLeu	486

```
Db 442 rlysaanserserThrThrTyrSerAlaLuhIsGluProGluSerGlyTyrLysThrTh 462
      ::::: |||::|
Qy 1276 AGGAGCTGTGTATGAGAACTCCAGACAG---GAGACCTTTCAGAGCAGCCCCAC 1332
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 462 rglY-SerAspTyrGlnGlnAlaValSerGlnArgGlnLagLutYrGluProGluThrV 482
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1333 TGGTGACAGAGCAGAGTGTGCTGTGACACATTACACACACATTACAGGCGAGC 1392
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 482 alTyrLnuValAlaGlyAlaGly-----AspHisTyrGlnLagLutGluAmet 498
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1393 TCAGTGGGCA-----GGGCTGTGCGCCGCTGCTGTACACTCCAG 1437
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 498 hTyrAspGluTyrGluAenGluLeuGlyLleThrAlaLleAlaLeuTyrAspTyrGlnA 518
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1438 CAGCCGACGACAGAGATCTCTTGCACCCGAGAACTCATCAAGGCGATCGAGTGA 1497
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 518 lalAlaGlyAspAspGluLleSerPheAspProAspAspLleLleThrAlaLleGluMet 538
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1498 TCAGCAGAGCTGTGCGCTGTATGGCGGAGTGCATTTTGGCATGTTCCCTGCCA 1557
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 538 lAspAspGlyTPTPTArgGlyVal---CysLysGlyArgTyrGlyLeuPheProAla 557
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1558 ACTACGTGAGCTC 1571
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 557 smTyrValGluLeu 561
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

RESULT 5
T02955
Probable cytochrome P450 monooxygenase - maize (fragment)

C/Species: Zea mays (maize)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #ext_change 23-Oct-1999
C/Accession: T02955
R/Feighlazar, K.; Strater, T.; Hachtel, W.
Submitted to the EMBL Data Library, March 1998
A/Description: Identification of a cytochrome P450-monooxygenase gene from corn.
A/Reference number: Z14790
A/Accession: T02955

A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-270 <YEG>
A/Cross-references: EMBL:AJ004810; NID:e1284328; PIDN:CAA06156.1; PID:e1284329

Alignment Scores:

Pred. No.: 9,49e-09 Length: 270
Score: 243.00 Matches: 66
Percent Similarity: 68.81% Conservative: 9
Best Local Similarity: 60.55% Mismatches: 23
Query Match: 5.75% Indels: 11
DB: 2 Gaps: 5

US-10-028-952a-4 (1-2333) x T02955 (1-270)

```
Qy 328 CGCCCGGTAACCTCCCACTGCTCCCGAGCGGCTCGCGCGCGCGCGCGC 269
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 3 ArgProSerGlnThrProHisLeuThrMetSerAlaArgLleGly-Pro-----Al 20
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 268 GCGGCGGCGGCTGTGGCGCAGAGCGAGAGCCCTCGGGGCTCGCGCGC---CGGCTC 212
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 20 aArgProGly-LeuGlyAlaLysArg-----GlyGlyAlaProLeuProThrH 37
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 211 ACCGGGTCAGTGAAGAAAGATCAGAGTACTGATTTTACCGCGCGCGCGCGCGCGG 152
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 37 lsgLylleSerLysLleThrLeuLysValAlaValPheHisLeuAArgProGluAlaPro 57
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 151 CGAGCCCGCGCGCGCGCGCTCGCGGAGACACCGCGGCGCGCGCGCGCGCGCAAGCTCA 92
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 hTyrProThrProLeuLysSer---PheHisLysValGlyLeuGluSer---SerSer 75
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 91 CAGGGTCTTCTTCCCGCTGATTCGCGCAAGCCGCTTCCCTGTGCTGTGCTTCCGCTG 32
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 75 hrgLysSerPheProAlaAspSerAlaLysProValProLeuAlaValaSerLeu 95
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
Qy 31 ATAGTAGTAGAGACAGTGGG 11
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 95 spSerArg-GlnAspSerGly 101
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

RESULT 6

Q083
BHLF1 protein - human herpesvirus 4 (strain B95-8)

C/Species: human herpesvirus 4, Epstein-Barr virus
C/Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #ext_change 23-Aug-1997
C/Accession: A03742
R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983

A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A/Reference number: A93065; MUID:85035713; PMID:6092825
A/Accession: A03742
A/Molecule type: DNA
A/Residues: 1-660 <BAN>

R/Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.;
Mature 310, 207-211, 1984
A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A/Reference number: A03794; MUID:84270667; PMID:6087149
A/Contents: annotation; protein coding region
A/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-5.

C/Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:

Pred. No.: 1.06e-08 Length: 660
Score: 242.50 Matches: 201
Percent Similarity: 30.67% Conservative: 41
Best Local Similarity: 25.48% Mismatches: 237
Query Match: 5.74% Indels: 310
DB: 1 Gaps: 43

US-10-028-952a-4 (1-2333) x Q083 (1-660)

```
Qy 2186 AGGTTGGAGAGTTGGCCATTCATCCCACTTCCCATGACC---CAGGCAGAGGACAG 2130
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 47 GlnValGlyLeuAlaAspAlaAlaSerProAspGluLeuInsPglAlaSerGlyAla 66
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 2129 AGAATGGAGTGGGAGACAGACAGAGCTCCAAAGAAATTCCTGTCGCAACC 2070
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 67 ArgProGlyGlyGlyAsn----- 72
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 2069 ACCATGATTCATCTGACTTGTGTCACAACTGCTAAACAAATTCCTACGTTACT 2010
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 73 -----Arg 73
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 2009 GTTCCCAAGGTCATCTTAAACAGTGTGGCCACACCCCTTAGAACAGAAAGAGA 1950
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 74 ValGlyAlaGlyArgGlyArgProGlyThrProAlaProSerArgInsArgArgThr 93
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1949 AAAAAGCAATTAATGACAGAAAGCAGACAGAGCTCCAGATGCCCTGCTCAGAACCC 1890
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 94 GlyProAlaGluGln-AlaSerHisAlaHisSerAsnProThrGlyLysSerAspPr 113
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1869 T-----TCCCGGAG-----CTTGGCAGTGGCCACTCAGGGCT 1857
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 113 cGlnArgSerProArgThrArgLlnAlaGlyTyrAlaLeuGlyGlyLysSerAlaGlyLe 133
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1856 TGGCCACAGCAGAGGCTTGGCTCAGTATCTGTGACAGGCGGACAGACAGCAAGCTGTGGGA 1797
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 133 uGlySerArg---GlyProArgProHisProAla-----PheGlnValGlnTrpSe 149
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1796 GGGTCCGGGAGATGACAGAGAGATGTGGGAATACACAGGCAATGTGCATTGGGGTG 1737
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 149 rAlaArg-----AsnProGlyCysProArgThrTTPArg-----ArgArgSerGlyAl 165
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1736 ACAGGCTAGTCTGACCAAGCCGAGAGGAGCCCTGAGGCTCAT---CCTCACTGGGGGT- 1661
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 165 aGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyPro-ThrGlyGlyA 185
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 1660 -----CCATTCCTGGAAGAGTGTGA 1659
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Db 185 rgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAla 205
 QY 1568 ATGCACTCCAGGCTCTCTCCAGCAATAGAGCCATGTCTGAGAGGGAGG 1599
 Db 205 laValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProPro 225
 QY 1598 CAAATGTGCTCCAGCTCACTCAATAGAGCTCCAGCTTGTGGCAGGGA 1549
 Db 225 laAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGluAspLeuAlaAla 245
 QY 1548 -----ACATGCCAAATAGG-----CATCGGCCCATAGCAAGCCACAGCCTT----- 1504
 Db 245 laGlnArgCysProAlaGlyProProThrArgSerGlyAlaAlaAlaGlnArgThr 265
 QY 1503 --CGTCATCACTCGATGCGCGGTGATGAGGTTCTCGGGGTCAAAGAGATCTCTGTG- 1447
 Db 265 laArgArgProProGlyCysPro-----ArgSerAlaArgAsnProGlyCysP 281
 QY 1446 -----CGTCGGCTGCTGTAGTCTGTACAGGGCAG----- 1417
 Db 281 roArgThrTPArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGly 301
 QY 1416 -----GGCAGCAGAGCCCTTGGCCCATGAGCCCTGAG----- 1384
 Db 301 laArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrPro 321
 QY 1383 -----CCTGAATGTGTGTCAATGTGTCTAGAGCCAGCAGCCTTGTCTGTCCAGCTG 1329
 Db 321 laAlaProGlyProGlyGlyGly-----AlaAlaValProSerG 334
 QY 1328 GAGCTGTCTGAGAGGTTCTCTCTGAGGTTCTCATACAGAGCTCTCTCTG 1269
 Db 334 lYAla-----ThrProHisProGluArgGlySerGlyP 345
 QY 1268 CTGCA-----CCAGACATGAGAGAGTGTGGGGCGCGCTCTCCAGCAGGAGATGTCCT 1212
 Db 345 roAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuPro 364
 QY 1211 GGGCTTGAAGATGAGCAGAGCTGCTCTGCGCAAGTGGGTCTGTGGTGGAGCTG 1152
 Db 365 -----GlnAspLeuAlaAlaGln----- 371
 QY 1151 CTTCTGAGAGAGGGCTCTCTGAGCTTCCAGGCTGAGAGCTGAGAGTGGAGTGT 1092
 Db 371 ----- 371
 QY 1091 GCGCCAGAGTCCCGCAGAGAGTGGCGTGGAGCGGCG-----GAGGGGG 1047
 Db 372 -----ArgCysProAlaGlyProProProThrArgSerGlyAla 384
 QY 1046 GCGCGCTTCCGCGCGCGCGCGCTTCCAGAG-----CGAAGGCACTCCGACCG 993
 Db 384 laAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProG 404
 QY 992 GA-----CCCCGGTCC-----CGCGCGCGCGG-----GGCGCGCGCGCTCCCGCG 951
 Db 404 lYCysProArgGlnThrTPArgArgArgSerGlyAlaGlnArgGlyHisProProPro 422
 QY 950 GCGCGCGCGCGCTGAGAGGGGGCGCGCGCGCGCGCGAGACAGCGGGGAGCCGCTA 891
 Db 423 -----GlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla 438
 QY 890 TCCGAGCGCAACGAGGCTCCGCGCGCGCTCCGATGTCTCCGCTGGGGGAGATTCTGA 831
 Db 438 laProGlyAlaPro-GlyThrProAlaPro----- 448
 QY 830 CTTAGAGCGCTTGCATATATCCACAGATGAGATTGCCCCATTTGCTCTCAGCC 771
 Db 449 -----GlyProGlyGlyGlyAla----- 454
 QY 770 AAGCACAATACCAATGTCTGAACCTGCGCTTCTCTCTGACTAGACAGATACCATG 711
 Db 455 -----AlaValPro-----SerGly 459

QY 710 GCACAACATCATCATAGAGGTAATAACTAGTCTCAGCAGCGTCTAATCCAGT 651
 Db 460 AlaThrProHisProGluArgGlySer----- 468
 QY 650 CACGTTCCCTATTAGTGGGTGAACATCAACAGCTTGGTGAATTCTCTCAATGATA 591
 Db 468 ----- 468
 QY 590 GGAAGAGCCGACATGAAAGATCAAAAAGCAGCTGCTATGAACGTTGGCCGACAA 531
 Db 469 GlyProAlaAsp-----ProProAla 475
 QY 530 GCCAGTTATCCCTGTGTACTTTTGTACACCTCTCTTTAAACCAAAAGTCAAGA 471
 Db 476 AlaAlaArg-----LeuProProGluArgGlnGlu 485
 QY 470 GGAATGTAGAGCCCGCTTTCACGCTGTGTCTGTAATCAAAATCAAGATCAAGCAGC 411
 Db 486 -----ProArgLeuPro-----GlnAspLeuAlaAla 494
 QY 410 TTT-----TGCCTTCTCTCTCAGCAGAGATTCTGTCTCTGAGCTGCTTGA 357
 Db 495 AlaGlnArgCysProAlaGlyPro-----ProProThrArgSerGly 508
 QY 356 CACCTGCGTTACCGTTTGAACAGTGTACG-----CCCACTCAAACTCCCACT 306
 Db 509 AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPro 528
 QY 305 GGCAGTGTCCCGGAGCGGAGTGCAGCGCGCGCGCGCG-----CCG 261
 Db 529 Gly--CysProArgThrTPArgArgArgSerGlyAlaGlnArgGlyHisProProProG 548
 QY 260 GCGCTTGGCGCCAGAGAGGAGAGCCCTCGGGGCTCGCC----- 220
 Db 548 lYAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProG 568
 QY 219 -----CCGCGCTCACCGGCTGAGTGAATAAAGATCAG-----AGTAGTGTATTTCACCG 168
 Db 568 lYThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaThrProH 588
 QY 167 CGGCGCGAGGGCGCGCGAGCCCGCGCGCGCTTGGCGGG-----ACACCGGGGGGCG 111
 Db 588 laProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlu 608
 QY 110 GCGGGGGGCGCAAGCTC 94
 Db 608 rgGlnGluProArgLeu 613

RESULT 7
 148899
 cortactin - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
 C:Accession: I48899; A49514
 R:Altarescu, M.R.; Wamton-Henderson, J.; Wu, H.; Parsons, J.T.; Bender, T.P.
 Oncogene 9, 1989-1997, 1994
 A:Title: The protein tyrosine kinase substrate cortactin is differentially expressed in
 A:Reference number: I48899; PMID:94268839; PMID:7510602
 A:Accession: I48899
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1546 <RBS>
 A:Cross-references: EMBL:U03184; NID:G414990; PIDN:AAA19689.1; PID:G509495
 R:Zhan, X.; Hu, X.; Hampton, B.; Burgess, W.H.; Friesel, R.; Maciag, T.
 J. Biol. Chem. 268, 24427-24431, 1993
 A:Title: Murine cortactin is phosphorylated in response to fibroblast growth factor-1 on
 A:Reference number: A49514; PMID:94043284; PMID:7693700
 A:Accession: A49514
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid; protein
 A:Residues: 1-8, 'R', 10-297, 'S', 299-546 <ZHA>
 A:Experimental source: BALB/c 3T3 cells

Query Match: 5.10% Indels: 326
 DB: 2 Gaps: 40
 US-10-028-952a-4 (1-2333) x J00405 (1-1106)

QY 2265 TTGAGGCGACACACTG-----GGGCGTGCAGCATCGAGTGAGCCGCTGGAGGAACATG 2212
 DB 126 LeuArgHisAspLeuArgArgGlyProAlaAlaLeuArgValAla-----Leu 142
 QY 2211 TTGTGTGCGCGTTTGAATACCAGGGTGGAGCTTGCCCATCTGCATCCCACTTCC 2152
 DB 143 LeuLeuArgPro----- 146
 QY 2151 CATAGCCGAGGCGACAGGAAGATGAGTGGGGAGACAGAGCAGGCTCAACAG 2092
 DB 147 HisValProGlyProGlyGlyAlaGlyArgGlyLeuHisArgGlyPro----- 163
 QY 2091 ACAAAATCCCTGCTGCCAAACAC---CATGATCCACTCTGACTTTGTCAAACTCTG 2035
 DB 164 -----ValProGlyArgValHisArgPro----- 171
 QY 2034 CTAAAAACATTTCTCACTGCTGCTCCCAAGGCTCATTTCAACAGTGTGCGCAC 1975
 DB 172 -----GluValHis-GlnProGlnProAlaLeuHisSerGlyAspHis 186
 QY 1974 AC----- 1973
 DB 186 HisArgSerLeuArgLeuHisAlaProAlaLeuGlyThrCysArgGlyValAlaLeuPro 206
 QY 1972 -----CCCTTGAGACAAAGAGAAAAAGCAAAATTAATG 1933
 DB 206 ValValArgArgAlaGlyGlyProAlaAspProAlaAspArgGlyProAlaArgGly 226
 QY 1932 CAGAGGCGACAGAGCTCCAGATGCGCCCTGAGAGCCTTCCCGCAG----- 1880
 DB 226 ValAlaArgAlaHisProLeuProGlyAlaArgHisArgGlyProArgProGlnGlyArg 246
 QY 1879 -----CTTGCGCATGGCCACTCAGGGCTTGCCACAGGCGCTTGCGCTC 1834
 DB 246 AlaArgGlyProValGlnGlyProValHisAlaGlyLeu--ArgAspArgGlyArgGly 265
 QY 1833 AGATACCTGTCAGGGCGAAGCCAAAGCTGTGGAGGTCGGGGATGCGAGGAAGAT 1774
 DB 265 sParArgProAlaLeuGlyProAlaArgAlaGlyAlaGlyGlnAlaHisArgArg 285
 QY 1773 GTGTGGAAATCACACAGCCCATTCGTCATTTGGGTGACAGGCTGAGTGCACAGCCG 1714
 DB 285 rArgArgGly-----ProProArgHisGlyGlyGlnHisProProAlaProHisArg 302
 QY 1713 AGGGAGCGCCGAGGCGCTACCTGCGGGGTCTATTCCTGGAGAGTGGCATGTC 1654
 DB 302 euGlyGlyAspArgProGlyAlaGlyGlyArgProArgArgGlyValArgGlyArg 321
 QY 1653 AACTCCCAAGGCTCTCTTCACCAAT----- 1627
 DB 322 GlValAlaArgArgGlyGlyGlnGlnHisArgGlyValArgArgAlaGlyArgArg 341
 QY 1626 -----AAGAAGCCATGTCT 1612
 DB 342 GlValProProValProLeuValLeuArgGlyAlaLeuLeuProGlnArgAlaArgAla 361
 QY 1611 GAGAGGGGAAGGCAAGATGTGCTCAGGCTCACTCAA-----TGAAGTCA 1564
 DB 362 AspArgGlyArgAspArgAlaAlaLeuValLeuGlnGlnProValArgArgValPro 381
 QY 1563 CGAGTGTGGAGGGAACAATGCCAAATGCCATCCGCGCCCATGCGCAAGCCAGCCTT 1504
 DB 382 Arg-ValHisArgHisArgLeuProProAla-GlyGlyProGlyProAlaArgArgGly- 400
 QY 1503 CGTGCATCAGCTGCATGCGCGCTGATGAGGTTCTCGGAGGTCAAAAGAGATCTGTGCT 1444
 DB 401 ArgArgAlaVal-----ProAlaArgGlyArgArgArgAlaValAla 415

QY 1443 CGG-----CTGCGTGTAGTCTGTACAGGGCGACGGGCACAGA 1408
 DB 416 ArgGlnValHisLeuGlyLeuLeuAlaAlaArgAlaArgArgAlaGlyGlnGlyAspGly 435
 QY 1407 GCCCTTGCCCACTAGGCCCTTGCGCTGAATGTGTGTCATGTGCTCAGAGCCAGCAC 1348
 DB 436 LeuLeu-----ProGly-----His 440
 QY 1347 CTGTGCTGCACACAGTGGGGCTGCTGTGAAAGTCTCTGCTCTG----- 1300
 DB 441 ProValGlnGlyProAspGlyGlyGlyAlaArgArgArgProAlaArgGlnGlyLeuGln 460
 QY 1299 -----GAGTTCTCATACACAGCTCTCTTGTGCTGCACACCATGAGAGAG 1249
 DB 461 GlValGlyAspValProGlnProValArgProArgAlaProLeuHisHisGlyLeuArg 480
 QY 1248 TGTGGGCGCG-----GTTCTCAGCAGGGAAGTCTGCGCTTGAGATGG 1198
 DB 481 GlyArgHisProLeuArgAspAlaGlnAlaArgGlyAsp----- 493
 QY 1197 CAGAGCTGCGCTCTGCAAAAGTGGTCTGTGAGTGAGCTGCTTCGACAGAAG 1138
 DB 494 -----ArgValGlyArg 497
 QY 1137 GCGTCTCAGCTTSCCAGG---CTGAGGACTGAGATGA---GTTGTTTACGCGCAGG 1084
 DB 498 -ArgProArgAlaLeuArgValValHisAlaGlyAspProValProGlyValProArgGly 517
 QY 1083 TTCCCGACGAGCGCGTGTGTCGAGCGGAGGGGCGCGCGCTTCCGCGCGCG-- 1026
 DB 517 ProProGlnProHisGlyProGlnArgAlaArgGlyArgProValHisArgGlyArgHis 537
 QY 1025 -CCGTTTCCGACGACGAAGGCGAC-----TC 1000
 DB 537 sProProAlaHisAlaArgGlyHisGlyValLeuLeuGlyValAlaAlaAspGlyProGly 557
 QY 999 CGACCGGACCGCGCTCCGCGCGG----- 975
 DB 557 ValAlaAspArgGlyProGlyAlaGlnGlyAspProGlyProAlaGlyValProAlaGly 577
 QY 974 -----CGCGCGGG 967
 DB 577 ArgArgProArgValProGlnProArgAlaAlaGlyArgHisProLeuArgArgArgGly 597
 QY 966 C---ACGCGCTTCCGCGCGCGCGCGGCGGTGAGAGGGGGCGCGCGCGCGGG 910
 DB 597 ProAlaHisProProGlyHisHisArgArgLeu-ArgAlaGlyArgArgProLeuArgP 617
 QY 909 ACAGCGGGGACCGGCTATCCGAGGCCAACCGAGGCTCCGCGCGCTGCGGTATCGTTC 850
 DB 617 roArgArgAlaValHisArgProAlaProAlaGlyGlnPro-----P 631
 QY 849 CGCTGGCGGGAATTGACTTAAGAGGCTTCACTCAATATCCACAGATGGTACCTTCG 790
 DB 631 roProHisArgAspProPro-----A 638
 QY 789 CCCCATTGGCTCCGACGACAGACATACCAAAATGTCGAACCTGGGTCTCTCTGT 730
 DB 638 lProAlaGlyProArgGlnHisProHisArgArgAlaArgArg-GlyHisAspArg 657
 QY 729 ACTGAGCAGATTAACATGCGCAACAACATCATCAGTGAAGGTAAACTAACCTGTCTCA 670
 DB 658 ArgGlyGlyLeuAspArgGlyHisArgProSer----- 668
 QY 669 CGAGCGTAAATCCAGGCTCAAGTTCCTATTAGGGGTGAACAATCCAGCTTGAGTA 610
 DB 669 ---ArgGlyArgValArgArgArgGlyArgAlaLeuGlyLeuProGlyGlySer----- 685
 QY 609 ATTGCTTCAACAATGATGAGGAAGAGCCGACATGGAAGATCAAAAAGCGAGCTGCTAT 550
 DB 666 -----GlnGlyGlnHisAlaValArgHisArgArgLeuProLeuArgProProLeu 702
 QY 549 GAAGCTTGGCGCGCAAGCCAGTTATCCCTGTGTAACTTTCTGACACACTCTCTGCTT 490

```

Db      703 HistHgtglaaglaaAlaSerArgPro----- 712
Qy      489 AAAACCCAAAGATCAGAGATGCTGAGCGCCGCTTTCAGCTCTGATTCCTACTGA 430
Db      713 -----GlyGluGlyAla-----ArgAlaAs 719
Qy      429 AAATCAAGATCAAGGAGCTTTTGCCCTTCCTCCTCAGCGAGGTTCTGCTCCTCTGA 370
Db      719 P-----GlyProArgArgProGluGluGlnProGl 729
Qy      369 GCTCCGCTTAGGA-----CACTGCGTTTACCTTT 340
Db      729 uGlyArgLeuGlyProGluProAlaArgGlyProHisGlyArgAspGlyValArgLe 749
Qy      339 GACAGCTGATACCGCCCGCATCAACTCCCGCATCGCATCTGCTCCCGAGCGAGT----- 285
Db      749 uArgGlnValHisAlaAspGlnArgAspPro-----LeuGlnGlyProGluGlnPro 766
Qy      284 -----CGCGCCCGCGCGCG-----CGCGCCCGCGCGCTTGCGCG 250
Db      766 cAlaGlnArgArgGlnAlaArgAlaArgProAlaProValArgGly-GlyThrArgAlaP 786
Qy      249 CAGAGCGAGAGACCCCTCGGGGCG---TCGCCCCCGCGCTCAGCGGTCAGTGAATAAAC 193
Db      786 GATGlyGlnGlyGlyProArgGlyProGluProHisArgAlaHisAlaThrLeuGlnProA 806
Qy      192 GATCAGAGTATGATTTTACCGCGCGCGCGCGAGCGCGG---CGAGCCCCCGCGCG 136
Db      806 rghs-----LeuHisGlyArgValArgArgAspProGlnA 818
Qy      135 CCCCTCGCGGAGAC---ACCGGGGGGGCGCGCGCGCGCGCGCAAGCTCAAGAGTCTT 80
Db      818 lLeuArgGlyAspProArgGlyGluGlyProGlyLeuProHisArgProValLeuLeu 837

```

RESULT 10

unspecific monooxygenase (EC 1.14.14.1) - common tobacco
T02995
N:Alternate names: cytochrome P450 homolog TBP
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T02995
R:Sugiyama, M.; Sakaki, T.; Yabueaki, Y.; Ohkawa, H.
Biochim. Biophys. Acta 1308, 231-240, 1996
A>Title: Cloning and expression in Escherichia coli and Saccharomyces cerevisiae of a nd
A:Reference number: Z14816; PMID:96404975; PMID:8809115
A:Accession: T02995
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-530 <SUG>
A:Cross-references: EMBL: D64052; NID: g1545804; PIDN: BAA10929.1; PID: g1545805
A:Experimental source: strain Bright Yellow 2
C:Genetics:
A:Gene: CTBP
C:Keywords: monooxygenase, oxidoreductase

Alignment Scores:
Pred. No.: 9.39e-06 Length: 530
Score: 199.00 Matches: 57
Percent Similarity: 61.1% Conservative: 9
Best Local Similarity: 52.78% Mismatches: 31
Query Match: 4.71% Indels: 12
DB: Gaps: 4

```

US-10-028-952a-4 (1-2333) x T02995 (1-530)
Qy      328 CGCCCAAGTCAAACTCCCACTGCACTGCTCCCGAGCGAGTTCGCGCGCGCGCGC 269
Db      3 ArgProSerGlnThrProHisLeuThrMetSerSerAlaArgLe----- 17
Qy      268 GCGCGCGCGCGCTTGCGCGCGAGAGAGCGCGCGCTTCGCGCGCGCGCGCGCT---- 213
Db      18 -----GlyLeuGlnHisAlaSerLeuGlySerLysLysArgGlySerAlaProLeuProIle 35

```

```

Qy      212 CACCGGCTCAGTGAATAAATCATCAGAGTATGATTTTACCGCGCGCGCGCGCGCG 153
Db      36 HisGlyIleSerLysIleThrLeuLysValValPheHisPheArgLeuSerAlaPro 55
Qy      152 GCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 93
Db      56 ThrTyThrThrProLeuLysSer---PheHisLysValGlyLeuGlnSer---SerSer 73
Qy      92 ACAGGCTCTTCTTCCCGCTGATTCGCGCAAGCCCGTTCCTTGCTGCTGCTGCTG 33
Db      74 ThrGlySerSerPheProAlaAspSerAlaLysProValProLeuValValSerLeu 93
Qy      32 GATAGTAGGTAGGACAGTGG 11
Db      94 AspSerArg-GlnAspSerGly 100

```

RESULT 11

ALR protein - human
T03455
C:Species: Homo sapiens (man)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 27-Oct-2003
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997
A>Title: Structure and expression pattern of human ALR, a novel gene with strong homology
A:Reference number: Z14954; PMID:97388474; PMID:9247308
A:Accession: T03455
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL: AF010404; NID: g2358286; PIDN: AAC51735.1; PID: g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: acute lymphoblastic leukemia protein, ALR type
C:Keywords: alternative splicing

Alignment Scores:
Pred. No.: 1.75e-05 Length: 4957
Score: 195.50 Matches: 178
Percent Similarity: 32.62% Conservative: 65
Best Local Similarity: 23.89% Mismatches: 271
Query Match: 4.62% Indels: 232
DB: Gaps: 33

```

US-10-028-952a-4 (1-2333) x T03455 (1-4957)
Qy      28 CTATCCAGGAAACACACAGCGAGGAGCGGCTTGCGGATCAGCG----- 75
Db      1546 lIeGlySerProThrThrPro-----AlaGlyLeuSerThrSerAlaAspGlyPheLeu 1563
Qy      76 -----GGAGAAAGAACCCCTGTTAGCTTGAGGCGCGCGCG 111
Db      1564 LysProProAlaGlySerValProGlyProAspSerProGlyGluLeuPheLeuLysLeu 1583
Qy      112 CCCCCCGGTGTCGCCCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 171
Db      1584 ProProGln-ValProAlaGlnAlaProSerGlnAspProPheGlyLeuAlaProAla-- 1602
Qy      172 TGAATACCACTACTGATCGTTTTCCTACTGACCGCGGTAGCGCGGCGG-----GGCG 225
Db      1603 ---TyProLeu-GluProArgPheProThrAlaProThrTyProProTyProS 1621
Qy      226 AGCCCGAGGCGCTGCTGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGA 285
Db      1621 erProThrGlyAlaProHisGlnProProMetLeuGlyAlaSerSerArgProGlyAlaG 1641
Qy      286 CCCGCTCGGGG-----ACAGTCCAGGTGGGAGATTGACTGGGCGCGGATCA 333
Db      1641 LysGln-ProGlyGluPheHisThrThrProProGlyThrProArgHisGlnProSerThr 1660
Qy      334 CCT-----GTCAACCGTAAACGAGGTGCTTAAAGCGAGCTCAGGAGAGACAGAA 384

```

Db	1661	ProAspProPheLeuLys-----ProArgCysProSerLeuAspAlaValPro	1678
Qy	385	ACCTCCCGCTGGAGCAGAAAGGGCAAAGCTGCCTGATCTTGATTTTCA-----	432
Db	1679	GlusErProGlyValGlyGlyGlyLysAlaSerGlnProLeuLeuSerProProPhe	1698
Qy	433	-----GTACAGATACAGACCCCTGAAAAGGGGCGCTCCAGATCCTTCT	474
Db	1699	GlyLusErArgLysAlaLeuGlnValLysLysGlnLeuGlnValAsaSerProSer	1718
Qy	475	-----GACCTTTTGGGTTTAAAGCAGAGAGGTGTCCAGAAAAGTTACACAGAGGATA	525
Db	1719	TyrGlyProProAsnLeuGlyPheValAspSer-----ProSerGly	1733
Qy	526	CTGCGCTTGGCGGCGCAAGCGTTATACAGCAGTCGCTTTTGATTCCTGATCGCGT	585
Db	1734	ThrIstLeuGlyGlyLeuGlnLeuLysThrProAspAlaPheLysAlaProLeuThrPro	1753
Qy	586	CTTCCTATCATTTGGAGCAGAAATTACCAAGCTGGATTGTTCCACCCTAATAAGGA	645
Db	1754	ArgAlaSerGlnValGlnProGlnSerProGlyLeuGlyLeu-----	1767
Qy	646	ACGAGAGTGGAGTATAGACCGCTCGAGACACGGTTAGTTACCCCTACGTATGATGT	705
Db	1768	-----ArgPro-GlnGlnPro-----ProProAlaGlnAlaLe	1778
Qy	706	GTTGCCATGTAACTCTGCTCAATACAGAGAGAAACCGAGGTTACAGACTTGGTGATG	765
Db	1778	UAlaPro-----	1780
Qy	766	TGCTTGGCTAGAGAACCAATGAGGGCGAAGTACACTGTGGATTATGACTGAACGCT	825
Db	1780	-----	1780
Qy	826	CTAAGTCAGAAATCCCGCCAGCGGAGACGATACGGACCGCGGAGCCTCGGTTGACC	885
Db	1781	-----SerProProSerHisProAspIlePheArgProGlySerTyrThrAspPr	1797
Qy	886	TCGATACCCGCTCCCGCGCGCTGTCCCGCGGCGGCGCGCGCC-----	931
Db	1797	o---TyrLaglnProPheLeuThrProArgProGlnProProProProGlnusErGy	1816
Qy	932	-----CCTCCACGCGCC---CGCGCGCGCGGAGGCGCGCTGCCCGCGCGCGCCG	981
Db	1816	salalauProProArgSerLeuProSerAspProPheSerArgValProValSerProGln	1836
Qy	982	GGACCGGGGCTCCGGTGGCGGAGTCGCCCTTCCTCTGGAAAGGGGCGCGGGAAGGC	1041
Db	1836	nserGlnErSerGlyGlnSerProLeuThrPro---ArgProLeuSerAlaGlnAl	1854
Qy	1042	GGCGCGCCCTTGGCCCTACGACGACCGGACGTTCTGTTGG---GAACTGGGCGTTAAACCA	1098
Db	1854	ahEcysProSerProValThr--ProArgPheGlnSerProAspProTylSerArgPro	1873
Qy	1099	CCTTCATCTCCAGT-----CTCAGCGTGGC	1124
Db	1874	ProSerArgProGlnSerArgAspProPheAlaProLysLysProProArgProGln	1893
Qy	1125	AAGCTGAGGAGCCCTTCTCTGAGAGGACGCTACCCACCAAGACCCACTTTGGCAGA	1184
Db	1894	ProProGlnValAlaPheLysAlaGlySerLeuAlaHisThrSerLeuGlyAlaGlyGly	1913
Qy	1185	GAGCCAGCTGCTGCCATCTCAAGGCGCCAGGAGCATCTCCCTCGTGAAGAGCGCGCGCC	1244
Db	1914	PheProAlaIleLeuProAlaGlyProAlaGlyGlnLeuHisAlaLysValaProSer---	1933
Qy	1245	AGCAGCTCTCATGTCTGTGACGAGAA-----	1274
Db	1933	GlyGlnProProAsnPheValArgSerProGlyThrGlyAlaPheValGlyThrProSer	1952
Qy	1275	-----GAGAGGCTGTGATGAG-----GAACCTCCAGAGGAC	1307

Db 1953 PromerAargPheThrPheProGlnAlaValGlyIuPProSerLeuTyPProValPro 1972

Qy 1308 GAGACCTTTACAGACAGACGCCCTACTGGTGCAGACGCAAGGCTCGGCTCGAGACATT 1367

Db 1973 Gln-ProGlyLeuProProProGlnIleGlyTLeuSerHisPheGlyProGlyProThrIle 1992

Qy 1368 G-----ACCAACCATTCAGAGGCGAGGCGCTC 1394

Db 1992 uGlyTyPProGlnSerThrAsnTyPThrValAlaThrGlyAsnPheHisPProSerGlySe 2012

Qy 1395 A--GTGGGMAAGGCTCT----- 1410

Db 2012 rProLeuGlyPProSerSerGlySerThrGlyGluSerTyGlyLeuSerProLeuArgPr 2032

Qy 1411 -----GTGGCCCTGCTGTACGACTACAGGACCGC-----AC 1445

Db 2032 oProSerValLeuProProProProAlaProAspGlySerLeuProTyLeuSerHisGlyAl 2052

Qy 1446 GACACAGAGATCTCCTTTGACCCCGAAGAACCTTCATCAGCGGAGATGAGGTATTCAGACGA 1505

Db 2052 aserGlnAArgSerGlyIleThrSerProValGluTyAsp-----GI 2066

Qy 1506 GGCTGGTGGCGGTGCTATGAGCGCGATGAGCATTTTGGCATGTCCTGCGCACTACG-- 1563

Db 2066 uAspProGlyThrGlyMetGly---SerSerLeuAlaThrAlaGluLeuProGlyTyThrG 2085

Qy 1564 -----TGAGCTCATTTGAGTGAGGCTGAGGACATCTTGCCCTCCCTCTGAG 1613

Db 2085 nAspProGlyMetSerGlyLeuSerGlnThrGluLeuGluTyGlnAArgGlnArgGlnhr 2105

Qy 1614 ACATGGCTTCCTTATT----- 1629

Db 2105 gLeuAArgIuLeuLeuIleArgGlnGlnIleGlnAArgAsnThrLeuArgGlnGluTyse 2125

Qy 1630 -----GCTGAAGAGAGAGGCGCTGGGAGTTGAC----- 1656

Db 2125 uThrAlaAlaAlaAlaAlaGlyAlaValGlyProProGlySerTyPProGlyAlaGluProse 2145

Qy 1657 -----ATTACAGACACTCTTCAGAAATAGAACCCCT-----AGTAGAGATGA 1697

Db 2145 rSerProAlaPheGlnGlnLeuSerArgGlyGlnThrProPheAlaGlyThrGlnAspIy 2165

Qy 1698 GGCCTCA-----GGGCTCCCTCCGGCTTGGCAGATCAGCCTGTCACCCCAATGCAAGC 1751

Db 2165 sSerSerLeuValGlyLeuProProSerTyLeuSerGlyProIleLeuGlyProGlyse 2185

Qy 1752 AATGAGCTGTGATTCCTCCACACATCTTCTGTCATCCCGACCCCTCCAGACAGCTTG 1810

Db 2185 r-PheProSerAspAspArgLeuSerArgProProProProAlaThrProSerSerMet 2204

RESULT 12

T03454

ALR protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #ext_change 27-Oct-2003

C:Accession: T03454

R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Onoegene 15, 549-560, 1997

A>Title: Structure and expression pattern of human ALR, a novel gene with strong homolog

A:Reference number: Z14954; MUID:97388474; PMID:9247308

A:Accession: T03454

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-5262 <PRA>

A:Cross-references: EMBL:AF010403; NID:G2358284; PIDN:AAC51734.1; PID:G2358285

C:Genetics:

A:Gene: ALR

A:Map position: 12

C:Superfamily: acute lymphoblastic leukemia protein, ALR type

C:Keywords: alternative splicing

Alignment Scores: 1.75e-05 Length: 5262

Db 2450 SerProAlaheGluGlnLeuSerArgGlyGlnThrProheAlaGlyThrGlnAspLeu 2470

Qy 1698 GGCTCA-----GGCTCCCTCGGGCTGGGACACACAGCTTCACACCCAAATGCAGC 1751

Db 2470 eSeSerLeuValaGlnLeuProSerLysLeuSerLysProLeuLeuGlyProGlySe 2490

Qy 1752 AATGCGCTGGTGATCCACACATCTTCTTCGATCCCGACCCCTCCGACAGCGCTTG 1810

Db 2490 r-PheProSerAspAspArgLeuSerLysArgProProProAlaThrProSerSerMet 2505


```

QY 549 CATAGCAGCTGGCTTTTGGATTCCTTGATGTGGCTCTTCTATCATTTGTGAAGCAGAA 608
Db 241 -----AspLeuAlaAlaGlnArg 247
QY 609 TTCACCAAGCGTTGATTGTTCACCACTAATAGGAGAGCTGAGGATTTAGCCGTC 668
Db 247 GCYSP-----AlaGlyProProProth 255
QY 669 GTGACAGGTAGTTTACCCTAGTANGATGTGTGTCAGTAGTAATCCTGCTCAG 728
Db 255 r----- 255
QY 729 TAGAGAGAGAACCGCAGGTTCAACATTTGGTATGTGCTTGGCTGAGAGCAATGGG 768
Db 255 ----- 255
QY 789 GCGAAGCTACCATCTGTGGATTATGACTGAAGCCTTAAGTCAGAAATCCGCCAGGC 848
Db 255 ----- 255
QY 849 GGAACGATACGGCAGAGCCCGCGAG-----CCTGGTTGGCTTCGA 890
Db 256 ---ArgSerGlyAlaAlaAlaGlnArgThnHisArgArgProProGlyCysProArgse 274
QY 891 TAGCCGCTCCCGCCCTGTCCTCCCGC----- 915
Db 274 rAlaArgAsnProGlyCysProArgThnTrpArgArgArgSerGlyAlaGlnArgGlyHn 294
QY 916 -----CGGCGGCGCGCCCGCCCGCTCCAGCGCGCCCGCGCGCG 953
Db 294 sProProProGlyAlaGlyGlnArgProSerGlyProThnGlyAlaArgProAlaAlaPr 314
QY 954 GGGAGGCGGCGTGGCCCGCGCGCGCGAGCGGGGTCGCGGAGAGCCCTTCGTC 1013
Db 314 oGlyAlaProGlyThnProAlaAlaProGlyProGlyGlyAlaAlaAlaValProBerG 334
QY 1014 CTGG-----GAAACGGGCGCGCGCGGAAAGCGCGCGCCCGCTCGCTCA 1061
Db 334 yAlaThnProHisProGlyArgGlySerGly---ProAlaAspProProAlaAlaAlaAr 353
QY 1062 CGCAGCCGACGTTCTGGGGAACCT----- 1086
Db 353 gLeuProProGlyArgGlnGlnProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCy 373
QY 1087 -GGCGCTAAACCACTTCATCTCCAGTCTCAGCC-----TGCAAGCTG 1130
Db 373 sProAlaGlyProProProThnArgSerGlyAlaAlaAlaGlnArgThnHisArgArgP 393
QY 1131 AGAAGCCCTTCCTCGCAGAAAGCAGCTCACCAACAGAGCCACTTTGGCAGAGAGCA 1190
Db 393 rProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThnTrpArgArg 413
QY 1191 GCTGCTCCATCTCAAGCCCAAGGCGAGATCTCCTGCTAGAGAGCGCGGCCAGCACT 1250
Db 413 er-----GlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProS 429
QY 1251 CCTCCATGCTGGGCA----- 1268
Db 429 eArgProThnGlyAlaArgProAlaAlaProGlyAlaProGlyThnProAlaAlaProG 449
QY 1269 GCAGAAAGAGAGGTGTGTATGAGAACTCCAGAGAGAGAGACTTTACAGCAGAGCC 1328
Db 449 lProProGlyGlyAla-----AlaValP 457
QY 1329 CCACTGTGTGACGACGCAAGGTGCTGAGTCTGAGACATTGACCAACCAAT----- 1378
Db 457 rSerGlyAlaThr-----ProHisProGlyArg 467
QY 1379 --TCAGGGCAGAGGAGCTCAGTGGCAAGGAGCTGTGCTCCG----- 1417
Db 467 lSerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlyArgGlnGlnProA 487
QY 1418 -----TGCCTGTA----- 1426

```

```

Db 487 rLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThnArgS 507
QY 1427 -----CGACTACAGGCGAG--CCGACGACAGAGAT 1456
Db 507 eArgAlaAlaAlaGlnArgThnHisArgArgProProGlyCysProArgSerAlaArgA 527
QY 1457 CTCCTTTACCCCGAAGACTCATCAGCGGATCG----- 1491
Db 527 snProGlyCysProArgThnTrpArgArgArgSerGlyAlaGlnArgGlyHisProProP 547
QY 1492 --AGGTATCGACAGAGCTGTGCGCTATAGGCGCGGATGAGCCATTGTGCAATGTT 1549
Db 547 roGlyAlaGlyGlnArgProBerGlyProThnGlyGlyAlaAlaValProBerGlyA 565
QY 1550 CCTGCCACTACGTGAGCTCATTTAGTGAAGCTGAGGAGCACTTGCCCTTCCTC 1609
Db 565 lAlaProGlyThnProAlaAlaProGlyProGlyGlyAlaAlaValProBerGlyA 585
QY 1610 TCAGACATGCTCTTATTGCTGGAAGAGAGCGCTGAGATTGACATTCAGCACTCTT 1669
Db 585 la-----ThnProHisProG 590
QY 1670 CCAGGATAGGACCCCGAGTGAAGATGAGGCTCAGGG--CTCCCTCCGGCTTGGCAGA 1726
Db 590 lArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlyArgGlnG 610
QY 1727 CTCAGCTGTGACCCCAATGACCAATGAGCTGTGATTCCACACATCTTCCTGAT 1786
Db 610 lProArgLeuProGlnAspLeuAlaAlaAla-----GlnArgCysProAlaG 626
QY 1787 CCCCCGACCTCCGACAGAGCTGTGCTCTTGGCCCTGACAGATAGTGAACCAAGCTTG 1846
Db 626 lPro---ProProThnArgSerGlyAlaAlaAlaGlnArgThnHisArgArgProBerG 645
QY 1847 CCTGT 1851
Db 645 lCys 646

RESULT 15
EDBE11
Immediate-early protein IE110 - human herpesvirus 1 (strain 17)
C.Species: human herpesvirus 1
C.Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 17-Mar-2000
C.Accession: A29152
R.Perry, L.J.; Rixon, F.J.; Everett, R.D.; Frame, M.C.; McGeoch, D.J.
J. Gen. Virol. 67, 2365-2380, 1986
A.Title: Characterization of the IE10 gene of herpes simplex virus type 1.
A.Reference number: A29152; MUID:87059760; PMID:3023529
A.Accession: A29152
A.Molecule type: DNA
A.Residues: 1-775 <PER>
A.Cross-References: GB:X04614; NID:G59832; PIDN:CA28285.1; PID:G59833
C.Genetics:
A.Introns: 19/3, 242/1
C.Superfamily: herpesvirus immediate-early protein IE110, RING finger homology
C.Keywords: DNA binding; early protein; transcription regulation; zinc finger
F.112-162/Domain: RING finger homology <RNG>
F.116-156/Region: zinc finger C3HC4 motif

Alignment Scores:
Pred. No.: 3,32e-05 Length: 775
Score: 191.00 Matches: 148
Percent Similarity: 31.07% Conservative: 49
Best Local Similarity: 23.34% Mismatches: 192
Query Match: 4,528 Indels: 246
DB: 1 Gaps: 30

US-10-028-952a-4 (1-2333) x EDBE11 (1-775)
QY 85 GACCGTGTGAGCTTGAGCGCGCGCGCGCGGTCGTCGCGGAGAGGCGCGGCGG 144
Db 240 AspAspAlaAspTyrValProProAlaProArgArgThnProAlaGlnAlaProProAlaArg 259

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 1, 2004, 15:58:27 Search time 100.5 seconds
(without alignments)
12178.976 Million cell updates/sec

Title: US-10-028-952a-4
Perfect score: 4230
Sequence: 1 gcacagagatccaccctgcgc.....agcttggaactcttcgcaca 2333

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1069545 seqs, 262320428 residues

Total number of hits satisfying chosen parameters: 2139090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xip
-Q/cgn2_1/USPTO.spool.p/US10028952a/runat.01042004.154628.775/app.query.fasta_1.2503
-Da=Published Applications_AA -QFMT=fastaln -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdt -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pico -NORM=ext -HEADSIZE=500 -MINDEN=0
-MAXLEN=2000000000 -USER=US10028952a@cgn.1.15@runat.01042004.154628.775
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG_SCORES=0 -WAIT -DEPLOC=100
-LONGLOC -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published Applications_AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	3664	86.6	793	14	US-10-028-952a-9	Sequence 9, Appl1
2	1480.5	35.0	784	14	US-10-028-952a-10	Sequence 10, Appl1
3	1191	28.2	525	15	US-10-084-886-114	Sequence 114, Appl
4	1189.5	28.1	355	9 <th>US-09-879-957-192</th> <th>Sequence 192, Appl</th>	US-09-879-957-192	Sequence 192, Appl
5	1024	24.2	270	12 <th>US-10-424-599-256311</th> <th>Sequence 256311,</th>	US-10-424-599-256311	Sequence 256311,
6	702.5	16.6	433	9 <th>US-09-879-957-18</th> <th>Sequence 18, Appl1</th>	US-09-879-957-18	Sequence 18, Appl1
7	476	11.3	90	9 <th>US-09-925-299-1477</th> <th>Sequence 1477, Appl</th>	US-09-925-299-1477	Sequence 1477, Appl
8	476	11.3	90	10 <th>US-09-925-299-1477</th> <th>Sequence 7840, Ap</th>	US-09-925-299-1477	Sequence 7840, Ap
9	472.5	11.2	118	14 <th>US-10-106-598-7840</th> <th>Sequence 1356, Ap</th>	US-10-106-598-7840	Sequence 1356, Ap
10	432	10.2	81	9 <th>US-09-925-299-1356</th> <th>Sequence 1356, Ap</th>	US-09-925-299-1356	Sequence 1356, Ap
11	432	10.2	81	10 <th>US-09-925-299-1356</th> <th>Sequence 1356, Ap</th>	US-09-925-299-1356	Sequence 1356, Ap
12	405	9.6	94	9 <th>US-09-925-299-1333</th> <th>Sequence 1333, Ap</th>	US-09-925-299-1333	Sequence 1333, Ap
13	405	9.6	94	10 <th>US-09-925-299-1333</th> <th>Sequence 1333, Ap</th>	US-09-925-299-1333	Sequence 1333, Ap
14	393	9.3	91	9 <th>US-09-925-299-1283</th> <th>Sequence 1283, Ap</th>	US-09-925-299-1283	Sequence 1283, Ap
15	393	9.3	91	10 <th>US-09-925-299-1283</th> <th>Sequence 1283, Ap</th>	US-09-925-299-1283	Sequence 1283, Ap
16	373	8.8	73	9 <th>US-09-925-299-1357</th> <th>Sequence 1357, Ap</th>	US-09-925-299-1357	Sequence 1357, Ap
17	373	8.8	73	10 <th>US-09-925-299-1357</th> <th>Sequence 1357, Ap</th>	US-09-925-299-1357	Sequence 1357, Ap
18	345	8.2	74	9 <th>US-09-925-301-1650</th> <th>Sequence 1650, Ap</th>	US-09-925-301-1650	Sequence 1650, Ap
19	332	7.8	74	9 <th>US-09-925-299-1402</th> <th>Sequence 1402, Ap</th>	US-09-925-299-1402	Sequence 1402, Ap
20	332	7.8	74	10 <th>US-09-925-299-1402</th> <th>Sequence 1402, Ap</th>	US-09-925-299-1402	Sequence 1402, Ap
21	324	7.7	58	9 <th>US-09-879-957-118</th> <th>Sequence 118, Appl</th>	US-09-879-957-118	Sequence 118, Appl
22	324	7.7	58	9 <th>US-09-879-957-118</th> <th>Sequence 214, Appl</th>	US-09-879-957-118	Sequence 214, Appl
23	309	7.3	59	10 <th>US-09-764-891-5255</th> <th>Sequence 5255, Ap</th>	US-09-764-891-5255	Sequence 5255, Ap
24	288.5	6.8	643	15 <th>US-10-369-493-5931</th> <th>Sequence 5931, Ap</th>	US-10-369-493-5931	Sequence 5931, Ap
25	270.5	6.4	19608	15 <th>US-10-084-846A-8</th> <th>Sequence 8, Appl1</th>	US-10-084-846A-8	Sequence 8, Appl1
26	267	6.3	54	9 <th>US-09-925-299-1314</th> <th>Sequence 1314, Ap</th>	US-09-925-299-1314	Sequence 1314, Ap
27	267	6.3	54	10 <th>US-09-925-299-1314</th> <th>Sequence 1314, Ap</th>	US-09-925-299-1314	Sequence 1314, Ap
28	266	6.3	54	10 <th>US-09-764-891-4667</th> <th>Sequence 4667, Ap</th>	US-09-764-891-4667	Sequence 4667, Ap
29	264.5	6.3	19652	15 <th>US-10-084-846A-7</th> <th>Sequence 7, Appl1</th>	US-10-084-846A-7	Sequence 7, Appl1
30	263	6.2	55	9 <th>US-09-925-299-1334</th> <th>Sequence 1334, Ap</th>	US-09-925-299-1334	Sequence 1334, Ap
31	263	6.2	55	10 <th>US-09-925-299-1334</th> <th>Sequence 1334, Ap</th>	US-09-925-299-1334	Sequence 1334, Ap
32	261.5	6.2	486	9 <th>US-09-879-957-46</th> <th>Sequence 26, Appl1</th>	US-09-879-957-46	Sequence 26, Appl1
33	259.5	6.1	19725	15 <th>US-10-084-846A-4</th> <th>Sequence 4, Appl1</th>	US-10-084-846A-4	Sequence 4, Appl1
34	258	6.1	91	14 <th>US-10-083-357-749</th> <th>Sequence 749, Appl</th>	US-10-083-357-749	Sequence 749, Appl
35	255	6.0	16695	15 <th>US-10-084-846A-3</th> <th>Sequence 3, Appl1</th>	US-10-084-846A-3	Sequence 3, Appl1
36	253	6.0	121	12 <th>US-10-424-599-221259</th> <th>Sequence 221259,</th>	US-10-424-599-221259	Sequence 221259,
37	251	5.9	19652	15 <th>US-10-084-846A-7</th> <th>Sequence 7, Appl1</th>	US-10-084-846A-7	Sequence 7, Appl1
38	250	5.9	763	15 <th>US-10-369-493-4212</th> <th>Sequence 4212, Ap</th>	US-10-369-493-4212	Sequence 4212, Ap
39	249	5.9	19723	15 <th>US-10-084-846A-5</th> <th>Sequence 5, Appl1</th>	US-10-084-846A-5	Sequence 5, Appl1
40	248	5.9	92	14 <th>US-10-083-357-750</th> <th>Sequence 750, Appl</th>	US-10-083-357-750	Sequence 750, Appl
41	246.5	5.8	19662	15 <th>US-10-084-846A-6</th> <th>Sequence 6, Appl1</th>	US-10-084-846A-6	Sequence 6, Appl1
42	242	5.7	19662	15 <th>US-10-084-846A-6</th> <th>Sequence 6, Appl1</th>	US-10-084-846A-6	Sequence 6, Appl1
43	238.5	5.6	19608	15 <th>US-10-084-846A-8</th> <th>Sequence 8, Appl1</th>	US-10-084-846A-8	Sequence 8, Appl1
44	231	5.5	546	9 <th>US-09-879-957-14</th> <th>Sequence 14, Appl1</th>	US-09-879-957-14	Sequence 14, Appl1
45	230	5.4	19695	15 <th>US-10-084-846A-3</th> <th>Sequence 3, Appl1</th>	US-10-084-846A-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-10-028-952a-9
Sequence 9, Application US/10028952a
Publication No. US20030157576A1
GENERAL INFORMATION:
APPLICANT: Evin, Jr., Paul R.
TITLE OF INVENTION: EPITHELIAL CELL GROWTH INHIBITORS
FILE REFERENCE: 4273.3USW1
CURRENT APPLICATION NUMBER: US/10/028,952A
CURRENT FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: PCT/US00/16900
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 60/139,995
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 793
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (16)..(17)
OTHER INFORMATION: Xaa is any amino acid or may be absent

```

1 FEATURE:
2 NAME/KEY: UNSURE
3 LOCATION: (42)
4 OTHER INFORMATION: Xaa is any amino acid or may be absent
5 FEATURE:
6 NAME/KEY: UNSURE
7 LOCATION: (45)
8 OTHER INFORMATION: Xaa is any amino acid or may be absent
9 FEATURE:
10 NAME/KEY: UNSURE
11 LOCATION: (49)
12 OTHER INFORMATION: Xaa is any amino acid or may be absent
13 FEATURE:
14 NAME/KEY: UNSURE
15 LOCATION: (65)..(66)
16 OTHER INFORMATION: Xaa is any amino acid or may be absent
17 FEATURE:
18 NAME/KEY: UNSURE
19 LOCATION: (82)
20 OTHER INFORMATION: Xaa is any amino acid or may be absent
21 FEATURE:
22 NAME/KEY: UNSURE
23 LOCATION: (99)
24 OTHER INFORMATION: Xaa is any amino acid or may be absent
25 FEATURE:
26 NAME/KEY: UNSURE
27 LOCATION: (115)..(116)
28 OTHER INFORMATION: Xaa is any amino acid or may be absent
29 FEATURE:
30 NAME/KEY: UNSURE
31 LOCATION: (122)
32 OTHER INFORMATION: Xaa is any amino acid or may be absent
33 FEATURE:
34 NAME/KEY: UNSURE
35 LOCATION: (132)
36 OTHER INFORMATION: Xaa is any amino acid or may be absent
37 FEATURE:
38 NAME/KEY: UNSURE
39 LOCATION: (136)
40 OTHER INFORMATION: Xaa is any amino acid or may be absent
41 FEATURE:
42 NAME/KEY: UNSURE
43 LOCATION: (149)
44 OTHER INFORMATION: Xaa is any amino acid or may be absent
45 FEATURE:
46 NAME/KEY: UNSURE
47 LOCATION: (165)..(166)
48 OTHER INFORMATION: Xaa is any amino acid or may be absent
49 FEATURE:
50 NAME/KEY: UNSURE
51 LOCATION: (174)
52 OTHER INFORMATION: Xaa is any amino acid or may be absent
53 FEATURE:
54 NAME/KEY: UNSURE
55 LOCATION: (179)
56 OTHER INFORMATION: Xaa is any amino acid or may be absent
57 FEATURE:
58 NAME/KEY: UNSURE
59 LOCATION: (182)
60 OTHER INFORMATION: Xaa is any amino acid or may be absent
61 FEATURE:
62 NAME/KEY: UNSURE
63 LOCATION: (199)
64 OTHER INFORMATION: Xaa is any amino acid or may be absent
65 FEATURE:
66 NAME/KEY: UNSURE
67 LOCATION: (205)
68 OTHER INFORMATION: Xaa is any amino acid or may be absent
69 FEATURE:
70 NAME/KEY: UNSURE
71 LOCATION: (215)..(216)
72 OTHER INFORMATION: Xaa is any amino acid or may be absent
73 FEATURE:

```

NAME/KEY: UNSURE	LOCATION: (232)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (236)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (249)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (265).. ₁ (266)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (274)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (282)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (287)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (299)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (313)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (315).. ₁ (316)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (332)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (349)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (365).. ₁ (366)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (380)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (382)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (399)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (415).. ₁ (416)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE	LOCATION: (432)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
NAME/KEY: UNSURE		

```

LOCATION: (449)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (465)..(466)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (482)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (499)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
LOCATION: (515)..(516)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (532)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (542)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (544)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

Alignment Scores:

```

Pred. No.: 1.76e-220 Length: 793
Score: 3664.00 Matches: 696
Percent Similarity: 87.77% Conservative: 0
Best Local Similarity: 87.77% Mismatches: 80
Query Match: 86.62% Indels: 17
Gaps: 1

```

US-10-028-952a-4 (1-233) x US-10-028-952a-9 (1-793)

```

QY 2 CACGAGATCCCATCTGTCCTACTATTCACGAAACCAACGCAAGGAAACGGGCT 61
Db 1 HistGluLeuProTyrValProThrTyrProAlaLysProGln*****GluArgAla 20
QY 62 TGGCGAATCAGCGGGAAGAAAGACCTGTAGC----- 97
Db 21 TrpArgAngLysGlyLysLeuThrLeuSerThrLeuValTrpHisGlyGlu 40
QY 98 -----TTGGGCCCCCGGGGCCCCCGCGGTTCGCCGAGG 133
Db 41 Thr**GluVal**AsnLysTrp**AlaProGlyAlaProProValSerProArgL 60
QY 134 GGGCGGGGGCGGGGCGCGGCGCGCGCGCGCGCGGTAATAACACTACTCTGATCG 193
Db 60 YAlaArgGlyLys*****ArgProGlyGlyProProValLysTyrHisTyrSerAsp 80
QY 194 TTTTTCACCTAGACCGGCGTGAAGCGGGGGGCGAGCCCGGAGGGGCTTCGCTTGGCGC 253
Db 80 Gphe**ThrAspProValArgArgGlyGlyGluProArgGlyAlaLeuAlaSer**Al 100
QY 254 CAAGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGGCGGCGGCGG 313
Db 100 alyArgTrpAlaAlaArgArgProGlyAlaThrHisSerGly*****AlaArgTrpG 120
QY 314 AGTTGACTGGGGCGGTAACACTGTCAACCGTAACGAGGTGTCTTAAGCGAGCTCAG 373
Db 120 yVal**LeuGlyArgTyrThrCysGlnThrVal**GlnValSer**GlyGluLeuAr 140
QY 374 GGAGGACGAAGAAACCTCCCGTGAAGCAAGGCAAAAGCTCGTGAATCTTGAATTTGAG 433

```

```

Db 140 gGluAspArgAnLeuProTyrSer**ArgAlaLysAlaArgLeuLeuLeuLeuPheSe 160
QY 434 TAGGAATACAGCCGTGAAGGGGGGCTCACGATCTCTTGACCTTTGGGTTTAAAGC 493
Db 160 rThrAsnThrAsp*****SerGlyAlaSerArgSerPhe**ProPheGlyPhe**Al 180
QY 494 AGGAGGTGCAGAAAGTTACACACAGGATTAAGTGGCTTGTGCGCGGCAAGGTTCAATG 553
Db 180 eGly***ValArgLysValThrThrGlyLLeThrClyLeuThrPArgProSerVal**Se 200
QY 554 CGAGCTCGCTTTTGAATCTTCGATGTGGGCTCTTCTTATCATGTTGTAAGCAATTCAC 613
Db 200 rAspValAlaPhe**SerPheAspValGlySerSerTyrHis*****AlaGluPheTh 220
QY 614 CAAGCGTTGATGTTTACCCCACTAATAGGGAACGTGAGGATTAAGCCGTCGTAG 673
Db 220 rlyArgTrpIleValHisProLeuIleGlyAsn**SerTrpAsp**ThrValValAr 240
QY 674 ACAGGTAGTTTACCTTACCTGATGATGATGTTTGCATGGAATTCCTGCTCAATACA 733
Db 240 gGlnValSerPheThrLeuLeuMet**CysCysCysHisGlyAsnProAlaGlnTyrG 260
QY 734 GAGGAACCGGACGATTCAGACATTTGTTGTTGCTGCTGAGAGGCAATGGGCGCAA 793
Db 260 uArgAsnArgArg*****HisLeuValTyrValLeuGly**GlyAlaAsnGlyAlaGly 280
QY 794 GCTACCATCTGTGGATTAATGACTGAACGCTCTTAAGTCAGAAATCCGCGGAGGAC 853
Db 280 sLeu**SerValGlyLeu**LeuAsnAlaSerLysSerGlnSerArgProGly**Th 300
QY 854 GATACGCGACGCGCGCGGAGCTCGGCTGCGGATACCGCGGCTCCGCTTCCCGC 913
Db 300 rIleArgGlnArgArgGlyAlaSerValGlyLeuGly**Pro*****ArgLeuSerPr 320
QY 914 GCGGCGGCGCGCGCGCGCGCTCCACGCGCGCGCGCGCGGCGGAGGCGCGTGCCTCC 973
Db 320 oProAlaGlyArgProProLeuHisAlaProArg**ArgGlyArgAlaArgAlaProPr 340
QY 974 GCGCGCGGAGCGGGGGTCCGTCGCGAGTGCCTTCTGCTGCGGAAACGGGCGCGCGC 1033
Db 340 oArgAlaGlyThrGlyValArgCys**ValProPheValLeuGlyAsnGlyAlaArgPr 360
QY 1034 GGAAGGCGGCGCGCGCGCGCTCGCGCTCAGCGACCCGACAGTTGCTGGGAAACCTGCGGCTA 1093
Db 360 oGluArgArgPro*****ArgProSerArgThrAlaArgSerTrpGlyThrTrpArg** 380
QY 1094 AACCACTTCATCTCAGTCTCAGCTCGGCAAGCTGAGAGAGCCCTTCTCAGAAAGCA 1153
Db 380 *Thr**SerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGln**G 400
QY 1154 GGTCAACCCAGACAGACCACTTTGGCAGAGAGCAGAGTGTGCATCTCAAGGCCGAG 1213
Db 400 nLeuThrGlnProGlnThrHisPheGlyArgGluProAlaAla*****SerArgProAr 420
QY 1214 GGCAGATCTCCTGCTGAGAGACCGCGCGCGCACTCTCATGTTCTGTGCGAGCAGA 1273
Db 420 gAlaAspLeuProAlaGlnGluProAlaProSer**ProProCysLeuValGlnAlaG 440
QY 1274 AGAGAGGCGTGTGTGAGAGAACTCCAGAGAGAGACCTTCTTCAGAGACGCCCACT 1333
Db 440 uGlnGluAlaValTyrGlnGluPro**GluGlnGlnThrPheTyrGlnGlnProProle 460
QY 1334 GGTGACGACGACGAGGTGCTGTGAGCACTTGAACATTAACACCAATTCAGAGGCGAGGCT 1393
Db 460 uValGlnGlnGln*****GlySerGlnHisIleAspHisIleGlnGlyGlnGlyLe 480
QY 1394 CAGTGGCGAAGGCGCTGTGCGCGCGCGCTGTAACGACTCCAGGAGCGGAGCGAGACAGA 1453
Db 480 uSer**GlnGlyLeuCysAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspAsp**G 500
QY 1454 GATCTCTTTGAACCCGAGAACCTCATCAACGGGCACTGAGGTGATGACGAAGGCGTGTG 1513
Db 500 uIleSerPheAspProGlnAsnLeuIleThrGlyIleGlnVal*****GluGlyTTrpTr 520

```

QY 1514 GCGTGGCTATGGGCGGATGGCATTTTGGCATTTTCCCTGCGCAATGAGTGGAGCTCAT 1573
D 520 pargglytrnglyproaspglyhisphglymet***ProLAsanTyrValGluLeu11 540
QY 1574 TGAGTAGGCTGAGGGGACATCTTGCCCTTCCCTTCAGACATAGGCTTCTTATGCTG 1633
D 540 eglu***Gly***GlyHisIleLeu***PheProSerGlnTrpLeuProTyrCysTr 560
QY 1634 GAAGAGAGGCTGAGGATTCATTCAGACACTCTTCAGAGATGAGAGAGAGAGAGAG 1693
D 560 plysargargpro*****HisSerIleLeuPheGlnGlu***AspProGln***G1 580
QY 1694 ATGAGGCTTCAGGGCTCCCTCCGGCTTGGAGACTAGGCTGTACCCCAATGAGCNA 1753
D 580 y****LeuArgIleProSerGlyLeuIleAspSerIleCysHisProLysCys***As 600
QY 1754 TGCGCTGGTATTCACACATCTTCCTGATCCCTCCGACACCTCCGACAGCTTGCT 1813
D 600 nsglyleuValIleProThrHisProSerCysIleProArgpro*****ThrIleTrpLe 620
QY 1814 CTTGCCCCCTGACAGATGACTGAGCCAGCCCTGCTGAGCCAGAGCCCTGAGTGGCCACT 1873
D 620 uleuProLeuThrGlyTyr***AlaLysProCys***TyrProSerProGluTrpProle 640
QY 1874 GCCAAGCTGGGGGAGAGGCTCCCTGAGCAGGAGGATCTGGGAGGCTTGCGCTGCTG 1933
D 640 uproserCysGlyGluGlySer*****GlyAlaSerGlyArgLeuTrpLeuProSer1 660
QY 1934 ATTTATTTGCTTTTCTTTTCTTTTCTTTGCTTAAGGGGTGGTGGCCACCACTGTTAG 1993
D 660 apheIleCysLeu*****PheSerLeuAlaSerLysGlyTrpTrpProProLeuPheArg 680
QY 1994 AATGACCCCTGGGAGACAGTGAAGTGAAGTGTGTTTGAAGAGCTTGTGACCAAGT 2053
D 680 gmet***LeuGlyAsnSerGlnArgGlnLeuPheLeuAlaGlnPheValThr***Va 700
QY 2054 CAGAGTGATCATGTGTGTTGGCAGCAGGAAATTTGCTTGTGGAGCTGCTGTCG 2113
D 700 lavgValAspHisGlyGlyLeuAlaIleGlyAsnLeuSerCys*****LeuLeuGlyAs1 720
QY 2114 TCCCATCTCATTTCTTCTGCTCCCTGCGCTGAGTGAAGTGGGAGTGGAGATGGCC 2173
D 720 apProHisSerIleSerLeuSerLeuGlyGly***GlyLysTrpGlyCysArgTrpTr 740
QY 2174 AAGCTCCACCTGGGTATTCAAAAGCGAGACACAACTGTCTTCCACGGGAGCTCAC 2233
D 740 oseSerHisProGlyLysSerLys***AlaAspThrThrCysSerSerThrArgLeuTh 760
QY 2234 TCGATGCTTCGAGGCCCCCAGTGTGCTCACTGATTTTCATGATTCAGGAAAGTAA 2293
D 760 rArgCysLeuGln*****ValCysAlaSerThrAspSerPheArgLysSerLys1 780
QY 2294 AAAAAAAAAAATCTGAGAGCTTTGAGCTTCTG 2330
D 780 slys***LysLysLeuGlnLysLeuTrpIleTrpIleSerSer 792

RESULT 2
US-10-028-952a-10
Sequence 10, Application US/10028952a
Publication No. US20030157576A1
GENERAL INFORMATION:
APPLICANT: Eryn, Jr., Paul R.
TITLE OF INVENTION: EPITHELIAL CELL GROWTH INHIBITORS
FILE REFERENCE: 4273,3USM1
CURRENT APPLICATION NUMBER: US/10/028, 952A
PRIORITY FILING DATE: 2001-12-18
PRIORITY FILING DATE: 2000-06-19
PRIORITY FILING DATE: 60/139,995
PRIORITY FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 10
LENGTH: 784
TYPE: PRN
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (110)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (13)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (16)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (33)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (49)..(50)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (66)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (83)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (90)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (99)..(100)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (104)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (116)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (133)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (142)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (147)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (149)..(150)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (166)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (173)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:


```

Db      118 ArgAlaLysAlaArgLeuIleLeuIlePheSerThrAsnThrAsp***GluSerGlyAla 137
QY      462 TCACGATCCTTGTGACCTTTGGGTTTAAGCAGAGAGGTGCAGAAAGTTACACAGAGG 521
Db      138 SerArgSerPhe**ProPheGlyPhe**Ala*****ValArgLysValThrThrGly 157
QY      522 ATAAGTGGCTGTGGCGGCGCAAGCTTCATAGCAGACGTGGCTTTTGAATCCTCGATGTC 581
Db      158 IleThrGlyLeuThrPArgProSer***HisSerAspValAlaPhe***SerPheAspVal 177
QY      582 GGCCTCTCTATCATTTGTAACAGAAATTCACCAAGCGTTGGATTGTTCACCCACTATA 641
Db      178 GlySerSerTyPHis**GluAlaGluPheThrLysArgTyrPLeValHisProLeuIle 197
QY      642 GGAACGTGAGCTGGAGATTAGACCGTCGTCGAGACAGTATTACCTTACCTGATGATG 701
Db      198 Gly*****SerTyrVal**ThrValValArgInValSerPheThrLeuLeu***Met 217
QY      702 TGTGTGTCATGATGTAATCTTCTCTCATGTAAGAGAGAAACCGCAGGTTCAAGCATTTG 761
Db      218 CysCysCysHisGlyAsnProAlaGlnTyrGluArgAsnArgArg**ArgHisLeuVal 237
QY      762 TATGAGCTGTGGAGGAGGAGCAATGGGGGAGAGCTACCATCTGTGGATTTGACTGAC 821
Db      238 TyrValLeuGly**GlyAlaAsnGlyAlaLys*****SerValGlyLeu***LeuAsn 257
QY      822 GCCTCTAAGTCAAGATCCCGCCGAGCGGAGCAAGATACGGCAGCGCCGCGAGCTCGATT 881
Db      258 AlaSerLysSerGlySerArgPro**GlyThrIleArgGlnArgArgGlyAlaSerVal 277
QY      882 GGCCTCGGATACCGCGGTCCCGCGCTGTCGCGCGGCGGCGGCGCGCC---CCCTCCAG 938
Db      278 GlyLeuGly**Pro**ProArgLeuSerProProAlaGlyArg-PropProPheSerTh 297
QY      939 GCGCCCGCGCGCGCGGAGGAGCGCGCTGCGCCCGCGCGCGCGGAGCGGGGTCCGGTGC 998
Db      297 Arg*****ArgAlaGlyGlyArgValProArgArgAlaProGlyProGlySer**Al 317
QY      999 GAGATGCCCTTCTGCTCTCGGGAACCGGGGCGCGGAGGCGGAGCGCGCCCTCGCCCG 1058
Db      317 aglucysproserSerTyrGlnThrGlyArgGlyArgLysGlyGly***ProLeuAlaAr 337
QY      1059 TCACGACCGCAGCTTCGT 1077
Db      337 gHisAlaProHisValArg 343

RESULT 3
US-10-094-886-114
/ Sequence 114, Application US/10094886
/ Publication No. US20040002120A1
/ GENERAL INFORMATION:
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Tchervnev, Velizar T.
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spytek, Kimberly A.
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Burgess, Catherine
/ APPLICANT: Vernet, Corinne A.
/ APPLICANT: Li, Li
/ APPLICANT: Gorman, Linda
/ APPLICANT: Malyankar, Uriel M.
/ APPLICANT: Boldog, Ferenc
/ APPLICANT: Guo, Xiaojia
/ APPLICANT: Shenoy, Suresh
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Taupier, Raymond J., Jr.
/ APPLICANT: Miller, Charles
/ APPLICANT: Casman, Stacie
/ APPLICANT: Pena, Carol
/ APPLICANT: Gangoli, Esha
/ APPLICANT: Gusev, Vladimir
/ APPLICANT: Smithson, Glenda
/ APPLICANT: Zerhusen, Bryan

```

```

/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Pochart, Pascal
/ APPLICANT: Fernandes, Elma
/ APPLICANT: Shimkels, Richard
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Spaderna, Steven
/ APPLICANT: LaRocheille, William
/ APPLICANT: Zhong, Mei
/ TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
/ FILE REFERENCE: 21402-290 B
/ CURRENT APPLICATION NUMBER: US/10/094,886
/ CURRENT FILING DATE: 2002-03-07
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/274,322
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/313,182
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/288,052
/ PRIOR FILING DATE: 2001-05-02
/ PRIOR APPLICATION NUMBER: 60/318,510
/ PRIOR FILING DATE: 2001-09-10
/ PRIOR APPLICATION NUMBER: 60/274,281
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/314,018
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: 60/274,194
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/274,849
/ PRIOR FILING DATE: 2001-03-09
/ PRIOR APPLICATION NUMBER: 60/296,693
/ PRIOR FILING DATE: 2001-06-07
/ PRIOR APPLICATION NUMBER: 60/313,626
/ PRIOR FILING DATE: 2001-08-21
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 298
/ SOFTWARE: Patentin 2.1
/ SEQ ID NO 114
/ LENGTH: 525
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-094-886-114

Alignment Scores:
Pred. No.: 3,57e-66 Length: 525
Score: 1191.00 Matches: 222
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 0
Query Match: 28.16% Indels: 1
DB: 15 Gaps: 0

US-10-028-952A-4 (1-2333) x US-10-094-886-114 (1-525)
QY      1095 ACCACCTTCATCTTCAGTCTCTAGCCTGCGCAAGCTGAGAGCCCTTCTTGCAGAACAG 1154
Db      304 ThrThrSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnLysGln 333
QY      1155 CTCACCCAGACAGAGACCCACTTTGGCAGAGAGCCAGCTGCTGCAATCTCAAGGCGCAGG 1214
Db      324 LeuThrGlnProGlnThrHisPheGlyArgGlnProAlaAlaAlaIleSerArgProArg 343
QY      1215 GCAGATCTCCCTGTGTGAGAGACCGGCGCCAGCACTCTTCATGTCTGTGACGACAGAA 1274
Db      344 AlaAspLeuProAlaGlnGluProAlaProSerThrProProCysLeuValGlnAlaGln 363
QY      1275 GAGAGAGCTGTGATATGAGAACTCCCGACAGACAGACCTTTCACAGAGAGCCGCCACTG 1334
Db      364 GluGluAlaValLysGluGluProGluGlnGlnThrPheTyrGluGlnProProLeu 383
QY      1335 GTGCAGACAGCAAGGTGCTGTGCTGTGACACATTGACCAATTCAGGCGCAGGCGCTC 1394
Db      384 ValGlnGlnGlnGlyAlaGlySerGluHisIleAspHisHisIleGlnGlnGlyLeu 403
QY      1395 AGTGGCAAGGCTCTGTGCGCCGCTGTAGACTACAGGAGCGAGCAGACAGAG 1454

```

Db 404 SerGlyGlnGlyLeuCySAAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspThrGlu 423
QY 1455 ATCTCCCTTACACCCGGAACCTCATCAGCGGCATCGAGTGAATGACGAAGGCTGTGG 1514
Db 424 ILSeSPheAspProGluMenLeuIleThrclyIleGluValIleAspGluGlyTPTP 443
QY 1515 CGTGGCTATAGGCGCGGATGGCCATTTGGCATGTTCCTTCGCAACTACGTGAGCTCAT 1574
Db 444 ArgGlyTyrGlyProAspGlyHisPhe-AlaCySerLeuProThrThrTyrSerSerle 463
QY 1575 GAGTGAAGCTGAGGAGGACATCTTCCCTTCCTTCAGACATGAGCTTCCTTATGCTGG 1634
Db 463 useGluAlaGlnGlyHisSerCysProSerProLeuArgHisGlyPheLeuIleAlaG 483
QY 1635 AAGAGAGGCTGAGGAGTGAATTCATTCAGACACTCTTCCAGAAATGAGACCCCGTAGGA 1694
Db 483 YArgGlyGlyLeuGlyValAspIleGlnHisSerSerArgAsnArgThrProSerGluAs 503
QY 1695 TGAGGCTCAGGAGGCTCCCTCCGCTTGGACAGCTCAGCTGTCACCCCAATGACCAAT 1754
Db 503 pGluAlaSerGlyLeuProProAlaTrpGlnThrGlnProValThrProAsnAlaAla 523
QY 1755 GGCGCTGG 1761
Db 523 LAlaTrp 525

RESULT 4
US-09-879-957-192
Sequence 192, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLER, Dana M.
MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pernie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 192:
US-09-879-957-192
Alignment Scores:
Pred. No.: 4,3e-66 Length: 355
Score: 1189.50 Matches: 251
Percent Similarity: 73.28% Conservative: 15
Best Local Similarity: 69.15% Mismatches: 63
Query Match: 28.12% Indels: 34
Gaps: 6
US-10-028-952A-4 (1-233) x US-09-879-957-192 (1-355)
QY 724 CTCAGTACGAGAGGAAACCGAGCTTCAGACATTTGGTGTAG-----TGCTTG 771
Db 12 LeuProGlyArgGlyThrProGlyProSerGlyLeuCyValProGluAspGlnCyArg 31
QY 772 GCTGAGAGCAATAGGCGGAAAGCTACATCTGGGATTAATGACTGAACCGCTTAAGT 831
Db 32 ValArgAspLeuLeuGlyTyrLeuAspSerPheThrAlaLysAlaGluLysGluGlu 51
QY 832 CAGAAATCCCGCCCAAGCGGAAAGTACGCGAGCCCGCGAGCCTCGGTTGCGCTCGAT 891
Db 52 AsnArgArgLeuGluGluLysArgTyrAlaGluGluAlaArgAlaGluGluGluGlu 71
QY 892 AGCGGCTCCCGCGCTCCCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
Db 72 ArgArgGluArgGluLeu-----ArgGluAlaAlaArgAlaGluGluGluGlu 89
QY 952 GCGGAGGAGC---GCGTCCCGCGCGCGCGC----- 979
Db 90 GlnGly-GlyGluAlaSerProGlnSerArgThrTyrGluGlnGlnGluValAla 109
QY 980 -CGGAGCGGCGTCCCGCGCGAGAGTCCGCTCGCTCGGAAACGCGGCGCGCGGAA 1038
Db 109 rArgAsnArgAsnGluGlnGluSerAlaValHisProArgGluLeu-PheLysGlnGlu 129
QY 1039 GCGGCG 1098
Db 129 LArg-----AlaSer-----ThrT 135
QY 1099 CCTCATCTCCAGATCTCAGCTGCGAGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCA 1158
Db 135 HisLeuSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnGlnLeuT 155
QY 1159 CCAACAGAGAGCCCACTTGGCAGAGAGCAGCTGCTGCATCTCAAGGCCAGGCGAG 1218
Db 155 HisGlnProGlnThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProAlaAla 175
QY 1219 ATCTCCCTGCTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1278
Db 175 sPheProAlaGluLeuProAlaProSerThrProProCysLeuValGlnAlaGluGlu 195
QY 1279 AGCTGTGATGAGAGACCTCCAGAGAGAGAGACCTTTCAGAGAGCGCGCGCGCGCTGC 1338
Db 195 LAlaValTyrGluGluLeuProProGluGlnGluThrPheTyrGluGlnProProLeuValG 215
QY 1339 AGCAGCAAGTCTGCTGCTGAGCAGCATTTGACCAACCAATTGAGGCCAGGCGCGCTCAGT 1398
Db 215 LngLngLngLysAlaGlySerGlnHisLeuPheHisIleGlnGlnGlnGlnGlnGln 235
QY 1399 GCGAAGGCTGTGCG 1458
Db 235 LngLngLysLeuCySAAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspThrGluIle 255
QY 1459 CTTTGAACCCGAGAACCTCATCAGCGGCGATGAGGTGATGACGAGAGGCTGTGGCTG 1518
Db 255 ePheAspProGluAsnLeuIleThrGlyIleGluValIleAspGlnGlyTPTPArgG 275
QY 1519 GCTATGGCGCGAGTGCATTTGAGCATGTTCCCGCAACTGAGTGAAGTCAATGAGT 1578
Db 275 LTyreGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIleAsp- 294

QY 1579 GAGGCTGAGGCGCATCTTCCCTCCCTCCAGATGAGCTTCTTATGCTGAGGA 1638
| | | | |
Db 295 GUAAGGUGLYTherCysProSerProLeuArgHisGlyPheLeuIleAlaGlyArg 314
QY 1639 GAGGCGCTGGAGATTACATTCAGCACTCTTCCAGAAATAGACCCCGATGAGATGAG 1698
| | | | |
Db 315 GYGLYLeuGlyValAspIleGlnHisSerSerArgAsnArgThrProSerGluAspGlu 334
QY 1699 GCGTCAGGCGCTCCCTCCGCGCTGGAGACTCAGCTCAGCCCAATCAGCAATGAGCC 1758
| | | | |
Db 335 AlSerGlyLeuProProAlaThrGlnThrGlnProValThrProAsnIleAlaMetAla 354
QY 1759 TGG 1761
| | |
Db 355 TTP 355
RESULT 5
US-10-424-599-256311
; Sequence 256311, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285664
; SEQ ID NO 256311
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(270)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_73471C.1.pep
US-10-424-599-256311
Alignment Scores:
Pred. No.: 8,93e-56 Length: 270
Score: 1024.00 Matches: 223
Percent Similarity: 82.42% Conservative: 2
Best Local Similarity: 81.68% Mismatches: 29
Query Match: 24.22% Indels: 21
Gaps: 6
US-10-028-952a-4 (1-2333) x US-10-424-599-256311 (1-270)
QY 1142 GAGGCGGCTCTCTCAGCTTGCAGAGCTGAGGACT-----GGAGAT 1104
| | | | |
Db 7 GUGGUYAIA-----AlaLeuLeuArgThrLysProArgProArgSerArgSer 23
QY 1103 GAGGCTGATTAGCGCCAGGCTTCCCAAGACGTGCGGTG---CGTACGCGGCGAGGAGG 1047
| | | | |
Db 24 ThrAsnGlyLeuAlaProGlySerThrArgThrCys-ValGlnArgAspGlyArgGluGlu 43
QY 1046 CGGCGCGCTTCCGCGCGCGCGCGCGCTTCCAGAGAGAGAGGAGCGCGAGCGGAGCCCG 987
| | | | |
Db 43 YGlyProLeuSerGlyArgThrProPheProLysThrAsnGlySerProHisArgThrPhe 63
QY 986 GGTCCCGCGCGCGCGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGGTGAGAGGAGG 927
| | | | |
Db 63 OValProThrProGlyValGlyArg-Arg-----ArgAlaIleAlaGlyThr-ThrA 80
QY 926 GCGGCG 870
| | | | |
Db 80 rGlyProArgArgAlaGlyThrAspGlyAspProAlaIleArgGlyGlnProArgLeu 100

QY 869 G-CGGCGCTCCGATGATTCCGCTGGCGGAGATTCTGACTTGAAGCGCTTGCATCA 811
| | | | |
Db 100 euaArgGlyArgGlyValProProGlyArgAspSerAspLeuGluAlaPheSerHisA 120
QY 810 ATCCCAAGATGATGCTTGGCCCATTTGGCTCTCTAGCCCAACACATTAACCAAAATGC 751
| | | | |
Db 120 snProThrAspGlySerPheAlaProLeuAlaProGlnProSerThrTyThrLysCysL 140
QY 750 TGAACCTGCGGTTCTCTCCACTAG-AGCAGATTACATGAGGAGAACACATCATCAT 692
| | | | |
Db 140 euaLeuAlaGlnPheLeuSerTy**SerArgIleThrMetAlaThrThiHisIleGln* 160
QY 691 AGGCTAAACTAACCTGCTCTCAGACGCTTAATCCAGCTCAGCTTCCCTATTAGGG 632
| | | | |
Db 160 **GlyLysThrAsnLeuSerHisAspGlyLeuAsnProAlaHisValProTy**Tpv 180
QY 631 TGAACAAATCCACGCTTGTGAATTCTGCTTCAATATGATAGAGAGCCAGATCGAG 572
| | | | |
Db 180 alaAsnProThrLeuGlyGluPheCysPheThrMetIleGlyArgAlaAspIleGluG 200
QY 571 GATCAAAAAGCGAGCTGCTATGAGCGCTTGGCCGCGCAGCAGCTTATCCCTGTGTA 512
| | | | |
Db 200 LysSerLysSerAspValAlaMetAsnAlaTrpProProGlnAlaSerTyProCysAla 220
QY 511 ACTTTCTGACACCTCTGCTTAAACCAACCAAGGTCA---GAAAGATCGTGAAGCCCG 455
| | | | |
Db 220 snPheSerAspThrSerCysLeuLysProLysArgSer**Arg-AspA-GluAlaPro 239
QY 454 CTTCACGCTGTATTGTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 395
| | | | |
Db 240 LeuSerArgSerValPheValLeuLysIleLysIleLysArgAlaPheAlaLeuLeu** 259
QY 394 CACGAGAGTTCTGCTCTCCCTCCGCTGAGCTCGCT 362
| | | | |
Db 260 HisGlyArgPheLeuSer***LeuSerSerPro 270
RESULT 6
US-09-879-957-18
; Sequence 18, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOWLES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mstock S Leelie
; REGISTRATION NUMBER: 18,872

Fri Apr 2 07:35:03 2004

us-10-028-952a-4.rapb

Page 10

PRIOR FILING DATE: 2000-03-08
 PRIOR APPLICATION NUMBER: 60/124,270
 PRIOR FILING DATE: 1999-03-12
 NUMBER OF SEQ ID NOS: 1556
 SOFTWARE: Patencin Ver. 2.0
 SEQ ID NO: 1477
 LENGTH: 90
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (53)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (76)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (87)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 LS-09-925-299-1477

Alignment scores:	
Pred. No.:	1,29e-21
Score:	476.00
Percent Similarity:	96.67%
Best Local Similarity:	96.67%
Query Match:	11.26%
DB:	10
Gaps:	0
Length:	90
Matches:	87
Conservative:	0
Mismatches:	3
Indels:	0
Gaps:	0

US-10-028-952A-4 (1-2333) X US-09-925-299-1477 (1-90)

DY 997 CACCGGACCCCGGTCCGGCCGCCGCAGCGCCCTTCCCGCGCGCGGAGCGC 938
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 HiskgThirProValProHlaArgIglYlaArgAlaleuProHrgAlaArgGlyAla 20
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 937 TGAAGGAGGGGGCGGCCCGCCGGCGGAGACAGCGCGGGGACCCTTATCCGAGGCCAAC 878
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 21 TTPrArgGLyGLyArGPrAlaGLyGLyAspArGArgGLyThnGLyTyPzProArGPrOthr 40

Dy 877 GAGCCTCCGCGCGCTACCGTATCTCCGCCGTGGCGGATTTCTGAATTAAAGCGCTTC 818
|||
Db 41 GIuaIaProIrgArgCysargIleValProIrogly***ApSerAapLeuAlaIaph 60

QY 817 AGCATTAATCCACAGATGTAGCTTCGCCCAATTGGCTCCTGAGCCAAGCACATACACC 758
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 SerHisAsnProThrAspGlnSerPheIleAlaProLeuIleAProGin***SerThirTyThr 80

QY 757 AATGTCGAACCTGGGTTCTCTGTAC 728
|||||
|||.|||||
Db 81 LysCysLeuAsnLeuArg**LeuSerTyr 90

RESULT 9
US-10-106-698-7840

; Sequence 7840, Application US/10106698
; Publication No. US20030109690A1

```

: *****
: GENERAL INFORMATION:
: *****
: APPLICANT: Ruben et al.

```

FILE REFERENCE: PA005PI
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27

CONVENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28

PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/157,137

PRIOR FILING DATE: 1999-11-01

```

; PRIOR FILING DATE: 1995-11
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0

```

```

; SOFTWARE: fad
; SEQ ID NO 7840
; LENGTH: 118

```

```

/ LENGTH: 110
/ TYPE: PRT
: ORGANISM: Homo sapiens

```

ORGANIZATION: AMERICAN SOCIETY OF MECHANICAL ENGINEERS

```
!
FEATURE: MISC_FEATURE
NAME/KEY:
POSITION: 450
```

LOCATION: (207)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE

```

; LOCATION: (61)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE

```

```

;
; LOCATION (b2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
;

```

```

1 LOCATION: (78)
2 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
3 NAME/KEY: MISC FEATURE
4

```

```

; LOCATION (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE

```

LOCATION: (82) _____
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE

LOCATION: (101)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids.

NAME/ACID: MISC. ESTERONE
LOCATION: (106)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

```

;
; NAME/KEY: MISC_FEATURE
;
; LOCATION: (118)
;
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

```

US-10-106-698-7840

Alignment Scores:

Pred. No.:	2.17e-21	Length:	118
Score:	472.50	Matches:	95
Percent Similarity:	87.16%	Conservative:	0

Best Local Similarity:	87.16%	Mismatches:	13
Query Match:	11.18%	Indels:	1
Pr:	14	Gaps:	1

US-10-028-952A-4 (1-2333) X US-10-106-698-7840 (1-118)

431 GAAATCAAGTCAAGCGAGCTTTGGCCCTTCTGCTCCACGGGAGGTTTGTCTCCACCT 372

371 GAGCTGGCCTTAGACACCTGCGTTACCGTTGACAGGTGTACCGCCCCAGTCAAACTCC 312

Db 29 GltLeuAlaLeuGlyHisLeuArgGlyTyrArgLeuThrGlyValProProGlnSerAsnSer 48

49 ProProGlyThrValProGlyAlaGly**AlaArg*****ArgGlyArgAlaLeuGly 68

251 GCCAGAGGAGAGCCCTCGGGCTCGCCCCCGGCTCACCGGGTCAGTGAATAACG 19

191 ATCAGAGTAGTGTATTTCACCGGCGGCCCCGCGAGCCCGCCCGGCGC--- 135
 Db 69 AlaargsergluserProserglylen**PROFIO*****Argvalserglulyslm 68
 Oy

Db 89 ILEI¹RGVALVALVALPHEH¹SAIRGARGPROALA¹RG*****LYSPROGLY***GLYALA 100

QY 134 CCCTCGCGGGACACCGGGGGCGGC 108
|||||
109 ProSerArrgIvIvSArGgIvG|vAla 117
|||||

RESULT 10
HE 00 02E 300-12EE

; Sequence 1356, Application US/09925299
; Patent No. US20020055627A1
; Patent Application

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1356
LENGTH: 81
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1356

Alignment Scores:

Pred. No.:	7,1e-19	Length:	81
Score:	432.00	Matches:	80
Percent Similarity:	98.77%	Conservative:	0
Best Local Similarity:	98.77%	Mismatches:	1
Query Match:	10.21%	Indels:	0
DB:	9	Gaps:	0

US-10-028-952a-4 (1-2333) x US-09-925-299-1356 (1-81)

QY 206 CCGGTGAGCGGGGGGCGAGCCCGGAGGGGCTCTGCTTGGCCGCAAGCCCGGC 265
DB 1 ProGlyGluAlaGlyGlyArgAlaProArgGlySerArgPheTrpArgIn***ProGly 20
QY 266 CCGCGCGCGCGCGGGCGGCGACCCGCTCCGGGACAGTGCAGGTCGGGGAGTTTGACTGGG 325
DB 21 ArgAlaProAlaGlyArgAspProLeuArgGlyGlnCysGlnValGlySerLeuThrGly 40
QY 326 GCGGTACCTGTCAACGGTAAAGCGAGTGTCTTAAGCGAGCTCAGGAGACAGAAA 385
DB 41 AlaValHisLeuSerAsnGlyAsnAlaGlyValLeuArgArgAlaGlnGlyGlnLys 60
QY 386 CCGCGGTGAGCAGAGGCGCAAAAGCTCGCTGATCTTGAATTTTGAAGATACAGA 445
DB 61 ProProValGluGlnLysGlyLysSerSerLeuAspLeuAspPheGlnTrpGluTrpArg 80
QY 446 CCG 448
DB 81 Pro 81

RESULT 11

US-09-925-299-1356
Sequence 1356, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1356
LENGTH: 81
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1356

Alignment Scores:

Pred. No.:	7,1e-19	Length:	81
Score:	432.00	Matches:	80
Percent Similarity:	98.77%	Conservative:	0
Best Local Similarity:	98.77%	Mismatches:	1
Query Match:	10.21%	Indels:	0
DB:	10	Gaps:	0

US-10-028-952a-4 (1-2333) x US-09-925-299-1356 (1-81)

QY 206 CCGGTGAGCGGGGGGCGAGCCCGGAGGGGCTCTGCTTGGCCGCAAGCCCGGC 265
DB 1 ProGlyGluAlaGlyGlyArgAlaProArgGlySerArgPheTrpArgIn***ProGly 20
QY 266 CCGCGCGCGCGCGGGCGGCGACCCGCTCCGGGACAGTGCAGGTCGGGGAGTTTGACTGGG 325
DB 21 ArgAlaProAlaGlyArgAspProLeuArgGlyGlnCysGlnValGlySerLeuThrGly 40
QY 326 GCGGTACCTGTCAACGGTAAAGCGAGTGTCTTAAGCGAGCTCAGGAGACAGAAA 385
DB 41 AlaValHisLeuSerAsnGlyAsnAlaGlyValLeuArgArgAlaGlnGlyGlnLys 60
QY 386 CCGCGGTGAGCAGAGGCGCAAAAGCTCGCTGATCTTGAATTTTGAAGATACAGA 445
DB 61 ProProValGluGlnLysGlyLysSerSerLeuAspLeuAspPheGlnTrpGluTrpArg 80
QY 446 CCG 448
DB 81 Pro 81

RESULT 12

US-09-925-299-1333
Sequence 1333, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1333
LENGTH: 94
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (15)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (38)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (42)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (88)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1333

Fri Apr 2 07:35:03 2004

us-10-028-952a-4.rapb

Page 13

Db 21 LeuAlaSerAspSerArgSerProAlaCysProArgArgAlaAlaProProSerThr 40
QY 940 CGCCCCCGCCGCGGAGGCGCCGTCGCCCCCGCCGCGGAGCCGGGATCCGATGCG 999
Db 41 ArgProAlaArgAlaGlyArgValProArgArgAlaProGlyProGlySerGlyAla 60
QY 1000 GAGTGCCCTTCGTCCTGGGAAACGCGGCGCGCGGAAAGCGCGCGC 1047
Db 61 GluCyProSerSerTrpGluThrGlyProGlyTrpLysGlyArg 76

RESULT 15
US-09-925-299-1283
Sequence 1283, Application US/09925299
Publication No. US2003040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1283
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (88)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1283

Alignment Scores:
Pred. No.: 1,95e-16 Length: 91
Score: 393.00 Matches: 74
Percent Similarity: 97.37% Conservative: 0
Best Local Similarity: 97.37% Mismatches: 2
Query Match: 9.29% Indels: 0
DB: 10 Gaps: 0

US-10-028-952a-4 (1-2333) x US-09-925-299-1283 (1-91)

QY 820 ACGCCTTAAGTCGATCCCGCCAGCGGAGATACGAGCCCGCGAGCCTCGG 879
Db 1 ThrProLeuSerGlnAsnProAlaGlnAlaGlyArgGlySerAlaAlaGluProArg 20
QY 880 TTGGCCTCGGATACCGCGGTCCTCCGCTGTCCTCCCGCGGCGCGCCCTCCACG 939
Db 21 LeuAlaSerAspSerArgSerProAlaCysProArgArgAlaAlaProProSerThr 40
QY 940 CGCCCCCGCCGCGGAGGCGCGGTGCCCCCGCCGCGGAGCCGGGATCCGGTGC 999
Db 41 ArgProAlaArgAlaGlyArgValProArgArgAlaProGlyProGlySerGlyAla 60
QY 1000 GAGTGCCCTTCGTCCTGGGAAACGCGGCGCGCGGAAAGCGCGCGC 1047
Db 61 GluCyProSerSerTrpGluThrGlyProGlyTrpLysGlyArg 76

Search completed: April 1, 2004, 16:17:15
Job time: 111.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 1, 2004, 15:46:31 ; Search time 108.5 Seconds

(without alignments)
12150.845 Million cell updates/sec

Title: US-10-028-952A-4

Perfect score: 4230

Sequence: 1 gcagcagatcccatcgtcc.....agcttgactcttcgcga 2333

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Command line parameters:

-MODEL=frame+ n2p -model: -DEV=xip
-Q=/cgn2_1/USPTO_epool_p/US10028952/runat_01042004_154624_634/app_query.fasta_1.2503
-DB=A -Geneseq_29Jan04 -QFMT=faelan -SUFFIX=rag -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTENT=ptc -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US10028952 @CGN 1.1 83 @runat 01042004_154624_634 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEIOBERRY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: A_Geneseq_29Jan04:*

1: geneseqp1980s:*\n2: geneseqp1980s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3389	80.1	760	4	AAB66394 Human pro
2	1370.5	32.4	753	4	AAB66395 Human mam
3	1191	28.2	525	6	ABUS5367 Human GPC
4	1189.5	28.1	355	2	AAW05398 Human CLO
5	1033	24.4	538	2	AAV23756 Human mam
6	872	20.6	302	5	ABR04724 Human pps
7	872	20.6	430	4	AAW78585 Human pro
8	870	20.6	458	4	AAW78569 Human pro
9	872	20.6	199	3	AAB65214 Human sec
10	870	20.6	203	5	ABB57455 Human sec

11	870	20.6	217	2	AAV73941 Human pro
12	870	20.6	377	6	ABP98850 Human str
13	870	20.6	431	3	AAB20896 Human dre
14	870	20.6	431	3	AAV85662 Human tyr
15	870	20.6	439	4	AAV93895 Human pro
16	716	16.9	162	2	AAV73942 Human pro
17	702.5	16.6	433	2	AAW05389 Mouse SH3
18	695	16.4	201	3	AAB63213 Gene 38 h
19	476	11.3	90	3	AAV53937 Human col
20	472.5	11.2	118	4	AAV7066 Human col
21	447.5	10.6	142	4	ABG14272 Novel hum
22	432	10.2	81	3	AAV53816 Human col
23	430	10.2	80	5	ABV57422 Human sec
24	430	10.2	80	6	ABP99731 Human sec
25	430	10.2	80	6	ABR01218 Human gen
26	430	10.2	80	6	ADA98322 Human sec
27	430	10.2	80	6	ADA44146 Human sec
28	430	10.2	80	7	ADC20492 Human sec
29	430	10.2	81	3	AAV63171 Human col
30	405	9.6	94	3	AAV53793 Human col
31	393	9.3	91	3	AAV53743 Human col
32	374.5	8.9	227	4	AAV33240 Novel hum
33	373	8.8	73	3	AAV53817 Human col
34	345	8.2	74	3	AAV44205 Human can
35	332	7.8	74	3	AAV53862 Human col
36	324	7.7	58	2	AAV34223 SH3 domai
37	324	7.7	58	2	AAV34248 SH3 domai
38	319	7.5	85	4	ABG00727 Novel hum
39	309	7.3	59	4	AAV95597 Human rep
40	308	7.3	301	4	AAV83949 Human imm
41	277	6.5	486	6	ABR44110 Human pro
42	277	6.5	486	6	ABO07214 Human p53
43	277	6.5	486	7	ADD67624 Human Ly1
44	267	6.3	54	3	AAV53774 Human col
45	266	6.3	54	4	AAV96009 Human rep

ALIGNMENTS

RESULT 1
AAB66394 standard: protein; 760 AA.

AAV73941

10-APR-2001 (first entry)

Human prostate ECGI protein sequence.

Human; mammatstatin A; mammatstatin B; mammatstatin C; ECGI; cancer; epithelial cell growth inhibitor.

Homo sapiens.

WO200078955-A1.

28-DEC-2000.

19-JUN-2000; 2000WO-US016900.

18-JUN-1999; 99US-0139995P.

(BIOT-) BIOTHERAPIES INC.

WPI; 2001-07193/08.

N-PSDB; AAF31282.

A mammatstatin-like epithelial cell growth inhibitor, and the nucleic that encodes it, useful for diagnosing and/or preventing epithelial cell

cancers, e.g. of the ovaries or prostate.

PS Disclosure: Page 41-43; 55pp; English.

XX The present invention describes a mammostatin-like epithelial cell growth
 CC inhibitor (ECGI) which has substantial similarity to mammostatin A, B or
 CC C. ECGI is expressed in healthy cells, but is either absent or expressed
 CC at reduced levels in cancerous cells. The protein and its coding sequence
 CC can be used to inhibit epithelial cell growth and the amount present in
 CC cells can be used to diagnose cancer or monitor its progress

XX Sequence 760 AA;

Alignment Scores:

Pred. No.:	4,13e-272	Length:	760
Score:	3389.00	Matches:	695
Percent Similarity:	87.64%	Conservative:	0
Best Local Similarity:	87.64%	Mismatches:	49
Query Match:	80.12%	Indels:	49
DB:	4	Gaps:	31

US-10-028-952A-4 (1-2333) x AAB66394 (1-760)

```

QY 2 CACGAGATCCCACTGTCCCTACTACTATCCAGGAAACCAAGCAAGGGAACGGGCT 61
DB 1 HSGIUILEProThValProThTyTyProAlaLysProGln****GluArgAla 20
QY 62 TGGCGGAATCAGCGGGGAAGAAGACCTGTGAGC----- 97
DB 21 TPARGAENGLNARGLYLysThrLeuLeuSerThrLeuValTrpHisGlyGlu 40
QY 98 -----TTGGGCCCCCGCGCGCCCGCGCGGCTGCTCCCGCGAGG 133
DB 41 Thr***GluVal***AsnLysTrp***AlaProGlyAlaProProValSerProArgG 60
QY 134 GGGCGCGGGCGGGGCGCGCGCGCTGCGGGCGCGGGTGAATCCACTACTCTGATCG 193
DB 60 YAlaArgGlyGly-----ArgProCysGlyProProValLysTyRHSIYSerAspAr 79
QY 194 TTTTTCACCTGACCCGAGGCGGGGGCGAGCGCCGAGGAGGCTCTGCTTGCGCGC 253
DB 79 G---PheThrAspProValArgArgGlyGlyGluProArgGlyAlaLeuAlaSer***Al 98
QY 254 CAAGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCTCCGCGGACAGTCCAGGTGGG 313
DB 98 aLysArgProAlaAlaArgArgProGlyAlaThrArgSerGly-----AlaArgTrpG 117
QY 314 AGTTGACTGGGGCGGTACACCTGTCAACGGTAACGACAGTGTCTTAAGCGAGCTCAG 373
DB 117 yVal***LeuGlyArgTyThrCysGlnThrVal---GlnValSer***GlyGluLeuAr 136
QY 374 GAGAGCAGAAACCTCCCGTGAAGCAGAAAGGCGAAAGCTCGCTGATGATTTTCA 433
DB 136 gGluAspArgAsnLeuProTrpSer***ArgAlaLysAlaArgLeuLeuLeuLeuPheSe 156
QY 434 TAGAATTAAGACCGGGAAGGCGGGCGCTCAAGATCTTCAACCTTTGGTTTAAAGC 493
DB 156 rTrpAsnTrpAsp-----SerGlyAlaSerArgSerPhe***ProPheGlyPhe---- 174
QY 494 AGGAGGTGCAGAAAGTTAACCAAGAGATAACTGGCTTTGGCGGCGCAAGCTTCATG 553
DB 174 *AlaGlyValArgLysValThrThrGlyLeuThrGlyLeuTrpArgProSerVal**Se 194
QY 554 CGAGCGCGCTTTTGAATCCGTGATGTCGGCTCTTCTATCATTTGGAAGCAATTCAC 613
DB 194 rAspValAlaPhe***SerPheAspValGlySerSerTyRHS-----AlaGluPheH 213
QY 614 CAAGCGTGTGATTTTCAACCACTAATAGGAACGTAGCTGGGATTAAGCTCGTGTAG 673
DB 213 rLysArgTrpIleValrHisProLeuIleGlyAsn---SerTrpAsp***ThrValValAr 232
QY 674 ACAGGTAGTTTAACTCTAGTATGATGTGTGTTCAGTGGTAATCCTGCTCAGATAGA 733
DB 232 gGlnValSerPheThrLeuLeuMet***CysCysCysHisGlyAsnProAlaGlnIlyrG 252

```

```

QY 734 GAGAACCCGACAGTTCAGACATTTGGTGTATGCTTGAGTGAAGCCAAATGGGGCGAA 793
DB 252 uArgAsnArgArg-----HisLeuValTyrValLeuGly***GlyAlaAsnGlyAlaTy 271
QY 794 GCTACCATCTGGGGATTATGACTGAAGCGCTCTAAGTGAATCCCGCCAGCGGAGAC 853
DB 271 sLeu---SerValGlyLeu***LeuAsnAlaSerLysSerGlySerArgProGly***Th 290
QY 854 GATACGACAGCGCGCGAGACCTCGGTTGGCTTCGAGATACCGGCTCCCGCTGTCC 913
DB 290 rIleArgGlnArgArgGlyAlaSerValGlyLeuGly***Pro***---ArgLeuSerPr 309
QY 914 GCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 973
DB 309 oProAlaGlyArgProProLeuHisAlaProArg--ArgGlyArgAlaArgAlaProPr 328
QY 974 GCGCGCGGAGACCGGGGTCGGGTGCGGAGTGCCTTCGTCTGGGAAACGGGCGCGGCC 1033
DB 328 oArgAlaGlyThrGlyValArgCys***ValProPheValLeuGlyAsnGlyAlaArgPr 348
QY 1034 GGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 1093
DB 348 oGluArgArg---Pro***ArgProSerArgThrAlaArgSerTrpGlyThrTrpArg-- 366
QY 1094 AACCACTCCATCTCCAGTCTCAGAGCTTGCAAGCTGAGAGAGCCCTTCCTGCAAGACA 1153
DB 367 ---TherIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGln***G 385
QY 1154 GCTCACCCAAACAGAGAACCACTTTGGCAGAGAGCGAGCTGTCTGCATCTCAAGCCG 1213
DB 385 nLeuTrpGlnProGlnThrHisPheGlyArgGluProAlaAla***---SerArgProAr 404
QY 1214 GCGAGATCTCTCTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1273
DB 404 gAlaAspLeuProAlaGluGluProAlaProSer---ProProCysLeuValGlnAlaG 423
QY 1274 AGAGAGGCTGTGTATGAGGAACCTCCAGAGCAGAGACCTTCTACAGACGCGCCACT 1333
DB 423 uGluGluAlaValTyGluGluPro**GluGlnGlnTrpPheTyGlnAlaAlaAspAsp**G 443
QY 1334 GGTGCAGCAGCAAGGTGCTGCTGAGCAGCATTTGACCAACCAATTCAGGGCGGGCT 1393
DB 443 uValGlnGlnGln---**GlySerGlnHisIleAspHisIleGlnGlyGlnIlyLe 462
QY 1394 CAGTGGCAAGGAGCTCTGTCCCGTGCCTGTACACTACAGGCGCGGAGCAKACACA 1453
DB 462 uSer---GlnGlyLeuCysAlaArgAlaLeuTyTrpArgTyGlnAlaAlaAspAsp**G 481
QY 1454 GATCTCTTGAACCCCGAGAACTCATCAGGGGATCGAGGTGATGACAGAGGCTGTG 1513
DB 481 uIleSerPheAspProGluAsnLeuIleThrGlyIleGluVal---**GluLysTrpTr 500
QY 1514 GCGGTGATGAGCGCGAGGAGCAATTTGGCATGTTCCCTGCCAATCTAGTGAAGTCA 1573
DB 500 pArgGlyTyGlyProAspGlyHisPheGlyMet---ProAlaAsnTyTrpAlaGluLeuI 519
QY 1574 TGGTGAAGGTGAGGGAGCAATCTGCGCTTCCCTCTCAGACATGAGTTCCTATTGCTG 1633
DB 519 eGlu***Gly***GlyHisIleLeu***PheProSerGlnThrTrpLeuProLysTr 539
QY 1634 GAAAGAGGCGCTGGAGTTGACATTCAGCACTTTCAGAGAAATAGAACCCCAAGTAGG 1693
DB 539 pLysArgArgPro*****HisSerAlaLeuPheGlnGln***AspProGln***Gl 558
QY 1694 ATGAGGCGCTCAGGGCTCCCTCGGCTTGGCAGACTAGCTGTCAACCCCAATAGCAAA 1753
DB 558 y---**LeuArgAlaProSerGlyLeuAlaAspSerAlaCysHisProLysCys***As 577
QY 1754 TGGCGTGTGATTTCCACACATCTTCTGCAATCCCGGACCTCCCAAGCAGCTGGCT 1813
DB 577 nGlyLeuValIleProThrHisProSerCysIleProArgPro-----ThrIleTrpLe 595
QY 1814 CTTCGCCCTGACGAGATACGAGCAAGCCAGCCCTGCTGCGCAAGCCCTGAGTGGCCACT 1873

```

Db 555 uleuPrleuthrthrglyYr***AlaIysProCys---TrpProSerProGluIutPrIole 614

QY 1874 GCCAAGCTGCGGGAGGAGGCTCTGAAGCAGGGGACATCTGGAGAGCTCTGGCTGCTTCTGC 1933

Db 614 uProSerCysglYgluIySer****GlyAlaSerIyArGlLeutPrIleuProSerAl 634

QY 1934 ATTATATTCCTTTTCTTTTCTTTTCTTTCTTTCTTTCAAGGGGAGGGTGGCCACACTGTATTAG 1993

Db 634 aPhelleCysIeu-----PheSerIleuAlaSerIyglYtrPrItrPrProIeuPheAr 653

QY 1994 AATGACCTTGGGAGCAGTGAACGAGAGAAATGTTTTTACAGAGTTTGGACCAAGT 2053

Db 653 gmet---IeuGlyAnSerGluAlaArgIleuPheIleuAlaGluPheValIu***Va 672

QY 2054 CAGAGTGATCATGTGTGTTTGGCAGCAGGAAATTTGTTTGTGGAGCTGCTCTGTGC 2113

Db 672 IArGValAspHisglYglYleuAlaAlaGlyAsnIeuSerCys----*LeuLeuCysAl 691

QY 2114 TCCCCACATCCATTTCTGTCTGTCCCTGTGCTGGGCTATGGAAAGTGGGATGAGATGGCC 2173

Db 691 aProHisSerIleSerIeuSerLeuCyIeuCly---GlyIystrIpgIyCysArgItrPr 710

QY 2174 AAGCTCCCAACCTGGGATATTCAAAACGGCAGACACAACAATGTTCTCTCAACGGGCTCAC 2233

Db 710 oSerIerHisProGlyIySerIyS**AlaAspThrIyCysSerSerItrArGlLeuth 730

QY 2234 TCGATGCTTGAGAGGCCACAGTGTGTGCTCTCAACTATTTCTGACTTCAGAGAAAGTAAAAA 2293

Db 730 rArgYsIeuGln****ValCysAlaSerThrAspSerAspPheArgIySer--Ly 748

QY 2294 AAAAAAAAAAAAAATCGAGAGAGCTTTGACATTTCTTCG 2330

Db 748 sIyAsIySlySlyIeuGluIySleutPrItrPheSer 760

CC	The present invention describes a mammostatin-like epithelial cell growth inhibitor (ECGI) which has substantial similarity to mammostatin A, B or C. ECGI is expressed in healthy cells, but is either absent or expressed
XX	
PS	Disclosure; Page 44-46; 55pp; English.
XX	
DR	WPI; 2001-071393/08.
XX	
PI	Ervin PR;
XX	
PA	(BIOT-) BIOTHERAPIES INC.
XX	
PR	18-JUN-1999; 99US-0139995P.
XX	
PB	
PF	19-JUN-2000; 2000WO-US016900.
XX	
PD	28-DEC-2000.
XX	
PN	WO200078955-A1.
XX	
OS	Homo sapiens.
XX	
KW	Human; mammostatin A; mammostatin B; mammostatin C; ECGI; cancer;
XX	epithelial cell growth inhibitor.
DE	Human mammostatin C protein sequence.
DT	10-APR-2001 (first entry)
AC	AAB66395;
ID	AAB66395 standard; protein; 753 AA.
RESULT 2	
AAB66395	

CC at reduced levels in cancerous cells. The protein and its coding sequence
CC can be used to inhibit epithelial cell growth and the amount present in
CC cells can be used to diagnose cancer or monitor its progress
XX
SQ Sequence 753 AA;

Alignment Scores:

Pred. No.:	1.14e-104	Length:	75
Score:	1370.50	Matches:	1
Percent Similarity:	88.04%	Conservative:	28
Best Local Similarity:	87.73%	Mismatches:	127
Query Match:	32.40%	Indels:	2
DB:	4	Gaps:	11

US-10-028-952A-4 (1-2333) X AAB66395 (1-753)

QY	102	CCCCCGGCGCCCCCGGATGCCCGCAGAGGGCCCGGGGCGGGATCCCGCGGCTGAC	161
Db	17	AlaProGlyAlaProProValSerProArgGlyAlaArgGlyGly***ArgArgProCys	36
QY	162	GGCGCCCGCGTGAAATACCACTACTCTGATCGTTTTTCTACTGACCCGGTGAGCGGGG	221
Db	37	GlyProProValAlaIstYrHisIstYrSerAspArg---***ThrAspProValAlaArgGly	55
QY	222	GGCAGAGCCCCGAGGAGGCTCGCTTCCTTCGGCGCCCAAGGCGCCCGGCGCGCGCCGGGC	281
Db	56	GlyGluProAlaArgGlyAlaLeuAla---GlyAlaIstArgProAlaAlaIstArgProGly	74
QY	282	GCGACCGGCTCCGGGGACAGTGCACAGGTGGGGAGTTGACTGGGGCGGTACCTGTCA	341
Db	75	AlaThrIstArgSerArgIst***SerAlaArgTrpAlaIst***LeuGlyArgGlyTrhCysGln	94
QY	342	ACGGTAAACGCGAGGTGCTCTAAGCGGAGCTCAGGAGACAGAAACCTCCCTGAGCAGA	401
Db	95	Thr---***GlnValSer***GlyGlnLeuArgGlnIstArgAlaLeuProTrp--Arg	112
QY	402	AGGCGAAAGTCCGCTTGATTTTATTTTTCACGTACGAAATACAGACCGTGAAAGGGGGCC	461
Db	113	ArgAlaIstYsAlaArgLeuIleLeuIlePheSerThrIstnIstAsp***GlnSerGlyAla	132
QY	462	TCAGATCCCTCTACCTTTGGSTTTTAAACAGAGAGGTGCAGAAAGTATCCACAGGG	521
Db	133	SerIstArgSerPhe***ProPheGlyPhe---***Ala***ValArgGlyValTrhTrhGly	151
QY	522	ATACTGGCTTGTGGCGGCGCAAGGTCATAGCGACGCGCTTTTTCATCTTGAGATGC	581
Db	152	IleThrGlyLeuTrpArgProSer--HisSerAspAlaAlaPhe***SerPheAspVal	170
QY	582	GGCTCTTCCTATCATTTGTGAAGCAGAAATTCACAGAGCGTTGGATTGTTCACCCACTAATA	641
Db	171	GlySerSerTrhIst***GlnAlaGlnPheTrhIstArgTrpIleValHisProLeuIle	190
QY	642	GGGAACGTGACCTGGAGATTACCCGTGAGAGAGTTATTTTACCTCACTGATATG	701
Db	191	Gly---***SerTrpVal**--ThnValValArgGlnValSerPheTrhLeuLeu--Met	208
QY	702	TGTTGTTCCTCATGTGTAATCTCTCTCAGTAGCAGAGGAAACGACAGTTTCAGACATTTGATG	761
Db	209	CysCysCysHisIstGlyAsnProAlaGlnTrpGlnArgAsnAspArg**ArgHisIleValAl	228
QY	762	TATGTCGTGCTGAGGAGCAATAGGGGCGAAGCTACCATCTGTGGATTTATGACTGATAC	821
Db	229	TyrValIstLeuGly***GlyAlaAsnGlyAlaIst***--SerValGlyLeu***LeuAsn	247
QY	822	GCCTTAAGTACAAATCCCGCCACAGCGGAAACGATACGGCAGCGCCCGCGGAGCTTCGGTT	881
Db	248	AlaSerIstYsArgGlnSerIstArgPro--GlyTrhIleArgGlnAspArgGlyAlaSerVal	266
QY	882	GAGCTCCGATACCGGATCCCGCGCTGCTCCCGCGCGGCGGCGCGCCCTCCACACG	941
Db	267	GlyIstLeuGly***Pro***ProArgIstLeuSerProAlaIstGlyArg-ProProProSerTh	286
QY	942	CCCCGCGCGCGGAGGCGCGTGGCCCCCGCGCGCGCGGAGACCGGGGTCGCGTGCAG	1001

Db 286 rArg**ArGAlaGlyGlyArGValProArGArGAlaProGlyProGlySer--AlaG1 305
 QY 1002 GTGCCCTTCCTCGTGGAAACGGGGCGCGCCGAAAGCGCGCCCTTCGCGCCGTCA 1061
 Db 305 uCysPProSerSerTertPgluThrGlyArgGlyArGlySerGlyGly***ProLeuAlaArgH1 325
 QY 1062 CGCACCGCACGCTTCCT 1077
 Db 325 sAlaProH1sValArg 330

RESULT 3
 ID ABUS2367
 XX ABUS2367 standard; protein; 525 AA.
 AC ABUS2367;
 XX 03-MAR-2003 (first entry)
 DT
 XX
 DE Human GPCR related protein NOV22a.
 XX
 KW Human; NOVX; G-protein coupled receptor; GPCR; cancer; cyrostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200279398-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 08-MAR-2002; 2002WO-US007355.
 XX

08-MAR-2001; 2001US-0274194P.
 PR 08-MAR-2001; 2001US-0274281P.
 PR 08-MAR-2001; 2001US-0274322P.
 PR 09-MAR-2001; 2001US-0274849P.
 PR 13-MAR-2001; 2001US-0275578P.
 PR 13-MAR-2001; 2001US-0275579P.
 PR 13-MAR-2001; 2001US-0275601P.
 PR 14-MAR-2001; 2001US-0276000P.
 PR 16-MAR-2001; 2001US-0276776P.
 PR 19-MAR-2001; 2001US-0276994P.
 PR 20-MAR-2001; 2001US-0277239P.
 PR 20-MAR-2001; 2001US-0277327P.
 PR 20-MAR-2001; 2001US-0277338P.
 PR 21-MAR-2001; 2001US-0277791P.
 PR 22-MAR-2001; 2001US-0277833P.
 PR 23-MAR-2001; 2001US-0278152P.
 PR 26-MAR-2001; 2001US-0278894P.
 PR 27-MAR-2001; 2001US-0278999P.
 PR 27-MAR-2001; 2001US-0279036P.
 PR 30-MAR-2001; 2001US-0280233P.
 PR 02-APR-2001; 2001US-0280802P.
 PR 02-MAY-2001; 2001US-0280852P.
 PR 02-MAY-2001; 2001US-0280866P.
 PR 02-MAY-2001; 2001US-0288228P.
 PR 17-MAY-2001; 2001US-0291766P.
 PR 07-JUN-2001; 2001US-0296693P.
 PR 08-JUN-2001; 2001US-0296866P.
 PR 05-JUL-2001; 2001US-0303230P.
 PR 05-JUL-2001; 2001US-0303237P.
 PR 08-AUG-2001; 2001US-0310913P.
 PR 13-AUG-2001; 2001US-0311978P.
 PR 14-AUG-2001; 2001US-0312191P.
 PR 16-AUG-2001; 2001US-0312916P.
 PR 17-AUG-2001; 2001US-0313182P.
 PR 20-AUG-2001; 2001US-0313408P.
 PR 21-AUG-2001; 2001US-0314018P.
 PR 27-AUG-2001; 2001US-0315227P.
 PR 10-SEP-2001; 2001US-0318403P.
 PR 10-SEP-2001; 2001US-0318510P.
 PR 14-SEP-2001; 2001US-0322296P.
 PR 14-SEP-2001; 2001US-0322360P.
 PR 27-SEP-2001; 2001US-0325378P.

PR 09-NOV-2001; 2001US-0332486P.
 PR 09-NOV-2001; 2001US-0345399P.
 PR 07-MAR-2002; 2002US-00094886.

(CURA-) CURAGEN CORP.

XX Kekuda R, Tchernev VN, Liu X, Spytek KA, Patunajan M;
 PI Burgess CE, Vernet CM, Li L, Gorman L, Malvaner UM, Boldog FL;
 PI Guo X, Shenoy SG, Padigaru M, Taupier RJ, Miller CE, Casman SJ;
 PI Pena CE, Gangoli EA, Gusev V, Smithson G, Zetrisen BD, Gerlach V;
 PI Pochart PF, Fernandes ER, Shinkets RA, Rastelli L, Spaderna SK;
 PI Larochele WJ, Zhong M, Karamtsov NV, Voss EZ, Herrmann JL;
 XX
 DR WPI; 2003-058423/05.
 DR N-PSDB; ABX70458.

PT NOVX polypeptides and polynucleotides, useful for treating a syndrome
 PT related to a human disease associated with the NOVX polypeptide e.g.,
 PT cancer.
 XX
 PS Claim 1; Page 156; 413pp; English.

CC The present invention relates to the isolation of novel human
 CC polypeptides referred to as NOVX (NOV1-NOV44), variants of these
 CC proteins, and the polynucleotide sequences encoding them. The NOVX
 CC proteins of the invention are G-protein coupled receptor (GPCR) related
 CC proteins. The sequences of the invention are useful in the manufacture of
 CC a medicament for treating a syndrome related to a human disease
 CC associated with the polypeptides e.g. cancer. ABUS2311-ABUS2408 represent
 CC the human NOVX proteins of the invention
 XX
 SQ Sequence 525 AA;

Alignment Scores:
 Pred. No.: 7.97e-90 Length: 525
 Score: 1191.00 Matches: 222
 Percent Similarity: 99.55% Conservative: 0
 Best Local Similarity: 99.55% Mismatches: 0
 Query Match: 28.16% Indels: 1
 DB: 6 Gaps: 0

US-10-028-952A-4 (1-2333) x ABUS2367 (1-525)

QY 1095 ACCACCTCCATCTCCAGTCTCCAGCCCTGAGAGAGAGCCCTTCTGACAGAGCAG 1154
 Db 304 ThrThrsertileserserProGlnProGlnLysleuArgerProPheleuGlnlyGln 323
 QY 1155 CTCACCCAGCAGAGAGACCACTTTGGCAGAGAGCCAGCTGCTGCATCTCAAGGCCAGG 1214
 Db 324 LeuthrglnProGlnuThrhAsPheGlyArGlnProAlaAlaAlaIleSerArGProArG 343
 QY 1215 GCATATCTCCCTGCTGAGAGAGCGGGCCAGACCTCTGCATGTGTGTGAGAGAGAGAA 1274
 Db 344 AlaAspLeuProAlaGlnGlnProAlaProSerThroProCysleuValGlnAlaGln 363
 QY 1275 GAGAGAGCTGTGTATGAGAGAACTTCAGAGAGAGAGACTTCTACAGAGAGCCGCCACTG 1334
 Db 364 GluGlnAlaValTyrglnGlnProProGlnGlnGlnThrPheTyrglnGlnProProleu 383
 QY 1335 GTGACAGAGAGAGGTGTGTGCTGAGACATTCAGACACATTCAGAGGCGAGGAGCTC 1394
 Db 384 ValGlnGlnGlnGlnAlaGlySerGlnAlaIleAspHisIleGlnGlnGlnGlnGln 403
 QY 1395 AGTGGAGAGAGGCTCTGTGCCGTGCGCTGTACACTTACAGAGAGAGAGAGAGAGAGAG 1454
 Db 404 SerGlyGlnGlyLeuCyAlaArGAlaLeuTyrsPyrglnAlaAlaAspAspThrgln 423
 QY 1455 ATCTCCTTTGACCCCGAGAGACCTCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
 Db 424 IleSerPheAspProGlnuAsnLeuIleThrglyIleGlnValIleAspGlnGlnTyrtip 443
 QY 1515 COTGACTATGGCGCGAGATGGCCATTTGGCATGTTCCCTGCCAATACGTGAGAGCTATT 1574

Db 444 ArgGlyTyrGlyProAspGlyHisPhe-AlaCysSerLeuProThrThrTrpSerSerLe 463
 QY 1575 GAGTGAAGGCTGAGAGGACATCTTGGCTTCCCTTCACACATGAGTCTTCTTATGCTGG 1634
 Db 463 uSerGluAlaGluGlyThrSerCysProSerProLeuArgHisGlyPheLeuIleAlaG 483
 QY 1635 AAGAGAGAGGCTGAGATTGACATTCAGCAGCTCTTCAGGAAATAGAGACCCCGAGTGAAGA 1694
 Db 483 yArgGlyGlyLeuGlyValAlaSplLeGlnHisSerSerArgAlaArgThrProSerGluAs 503
 QY 1695 TGAAGGCTCAGAGGCTCCCTCCGCTTGGCAGATCTCAGCTGTACCCCAATGCGAGCAT 1754
 Db 503 pGluAlaSerGlyLeuProProAlaTrpGlnThrGlnProValThrProAlaAlaIa 523
 QY 1755 GGCCTGG 1761
 Db 523 tAlaTrp 525
 RESULT 4
 AAM05398
 ID AAM05398 standard; protein; 355 AA.
 AC AAM05398;
 XX 19-FEB-1998 (first entry)
 XX Human clone 56 protein.
 DE
 XX Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
 KM cellular signalling element; cellular structural element; malignancy;
 KM protein identification; functional domain; protein screening;
 KM cellular signal transduction process.
 XX Homo sapiens.
 OS
 XX MO631625-A1.
 PN 10-OCT-1996.
 XX 04-APR-1996; 96WO-US004454.
 PF 07-APR-1995; 95US-00417872.
 PR 03-APR-1996; 96US-00630915.
 XX (CYTO-) CYTOGEN CORP.
 PA (UTNC-) UNIV NORTH CAROLINA.
 XX Sparks AB, Hoffman N, Kay BK, Fowlkes DM, McConnell SJ;
 DR WPI: 1996-465045/46.
 DR N-PSDB: AAT39798.
 XX
 PT Identifying polypeptide(s) having specific functional domain (esp. SH3
 PT domain) - comprises detecting selective binding to recognition unit,
 PT regardless of sequence homology.
 XX
 PS Claim 54; Fig 57; 174pp; English.
 XX
 CC AAM05386-M05403 represent novel human and mouse Src-homology region 3
 CC (SH3) domain containing proteins that can be used in the method of the
 CC invention. SH3 domain containing proteins play a role in signalling and
 CC structural elements of cells. The method of the invention is for
 CC identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). The method comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins which contain an SH3 domain due to the minimal sequence homology
 CC among known SH3 proteins. It has been found that small peptide RUs in
 CC multivalent form have reduced specificity for a given functional domain
 CC compared to monomer RUs. Multivalent RU complexes are particularly suited
 CC to screening for polypeptides containing functional domains that are

CC similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention
 CC
 XX
 SQ Sequence 355 AA;
 Alignment Scores:
 Pred. No.: 9,51e-90 Length: 355
 Score: 1189.50 Matches: 251
 Percent Similarity: 73.28% Conservative: 15
 Best Local Similarity: 69.15% Mismatches: 63
 Query Match: 28.12% Indels: 34
 DB: 2 Gaps: 6
 US-10-028-952A-4 (1-233) x AAM05398 (1-355)
 QY 724 CTCAGTACGAGAGAACCCGAGTTCAGACATTGGTGTATG-----TGCTTG 771
 Db 12 LeuProGlyArgGlyThrProGlyProSerGlyLeuGlyValProGluAspGlnCysArg 31
 QY 772 GCTGAGAGCCAAATGGGGCGAAGGCTACCATCTGTGGGATATGACTGAACGCTCTAAGT 831
 Db 32 ValArgAspLeuGlyGlyThrLeuAspSerPheThrAlaValAlaGluValGluGlu 51
 QY 832 CAGATCCCGCCAGCGGCGAAAGATACGAGCGCCCGCGAGCTCGTGGCTCGCAT 891
 Db 52 AsnArgArgLeuGluGluValArgTrpAlaGluAlaGlnArgGlnLeuGluGlu 71
 QY 892 AGCGGTCGCCCGCGCTGCCCCCGCGGCGCGCGCCCGCCCTCAGCGCGCCGCGCGC 951
 Db 72 ArgArgGluArgGluLeu-----ArgGluAlaAlaArgGluGlnArgTyrGlnGlu 89
 QY 952 GCGGAGAGGCG--GCGTGCCTCCCGCGCGCGC----- 979
 Db 90 GlnGly-GlyGluAlaSerProGlnSerArgThrTrpGluGlnGlnGluValAlaIse 109
 QY 980 -CGGACCGGAGTCCGCGCGGAGTCCCTCGTCCGCGGAAACGGGCGCGCGCGGAAA 1038
 Db 109 rArgAsnArgAsnGluGlnGluSerAlaValHisProArgGluIle-PheIysGlnIysG 129
 QY 1039 GCGGCGCGCGCGCGCGCGCTCAGCGACGACGTTGCTGGGAAACCTGGCGCTAAACCA 1098
 Db 129 lArg-----AlaMetSer-----ThrT 135
 QY 1099 CCTCATCTCCAGTCTCTTACGCTGAGGAGAGCCCTCTCTCGAGAGACGCTCA 1158
 Db 135 hSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuIleIysGlnLeu 155
 QY 1159 CCCAACCGAGAGCCACTTTGGCAGAGCGGAGTGTGGCATCTCAAGGCCCGAGGAG 1218
 Db 155 hGlnProGluThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProArgAla 175
 QY 1219 ATCTCCCTGCTGAGAGCGCGCGCGCCAGCACTCTCTCAATGTCTGTGGAGCGAAGAG 1278
 Db 175 sPleuProAlaGluGluProAlaProSerThrProCysLeuValIleAlaGluGlu 195
 QY 1279 AGCGTGTATGAGAGACCTCCAGAGCGAGAGACCTTACAGAGAGCGCCCACTGTGTC 1338
 Db 195 lAlaValIleGluGluProProGluGlnGluThrPheTyrGluGlnProProLeuValG 215
 QY 1339 AGCAGCAAGTCTGCTGTGAGCATTTGACCACTTACAGAGCGCGAGGCTCACTG 1398
 Db 215 lGlnGlnGlyAlaGlySerGluHisIleAspHisHisIleGlnIleGlnIleuSerG 235
 QY 1399 GCGAAGGCTGTGTCGCGCGTGTACGACTTACAGAGCGAGCGACGACGAGATCT 1458
 Db 235 lYglnGlyLeuGlyAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspAspThrGluIle 255
 QY 1459 CCTTGAACCCGAGAACCTCATCAGCGGCGATCGAGGTGATCGACGAGAGGCTGTGGCGTG 1518

```

Db      255 eRPheSPProGluSnLeuIleTrGlyIleGluValIleAspGluGlyTrpTrpArg 275
QY      1519 GCTATGGCGCGGATGGCCATTGTGGCATGTTCCCTGCCAATGCTGAGCTCATTAAGT 1578
QY      275 IYTyGlyProAspGlyHisPheGlyMetPheProAlaAsnTyValGluLeuIleAsp 294
QY      1579 GAGGCTGAGGCAACATTTGCCCTTCCCTCCAGACATGCTTCTTATTCCTGAGAGA 1638
QY      295 GUAAlaGluGlyHisSerCysProSerProLeuArgHisGlyPheLeuIleHisGlyArg 314
QY      1639 GAGGCGCTGAGATTGACATTCAGCACTCTTCAGAAATAGACCCCGAGTGAAGATGAG 1698
Db      315 GlyGlyLeuGlyValAlaSerIleGlnHisSerSerArgAsnArgThrProSerGluAspGlu 334
QY      1699 GCCTCAGGCGCTCCCTCCGCTTGGCAGATCAGCTGTACCCCAATGCGAATGGCC 1758
Db      335 AlaSerGlyLeuProProAlaTrpGlnThrGlnProValThrProAsnAlaIleVal 354
QY      1759 TGG 1761
Db      355 Trp 355

```

RESULT 5

AA23756 standard; protein; 538 AA.

AA23756;

09-SEP-1999 (first entry)

Human mammary epithelial cell growth inhibitor.

Human mammary epithelial cell growth inhibitor.

Human; mammary epithelial cell cancer.

Homo sapiens.

W0932625-A2.

01-JUL-1999.

18-DEC-1998; 98MO-US027147.

19-DEC-1997; 97US-00994076.

(UNMT) UNIV MICHIGAN.

Ervin PR;

WPI; 1998-418924/35.

N-PSDB; AA85833.

DNA encoding Mammary epithelial cell growth inhibitor.

Claim 2; Page 45-47; 77pp; English.

The present sequence represents human mammary epithelial cell growth inhibitor.

Mammary epithelial cell growth inhibitor in culture and recombinant expressed

CC by normal human mammary cells in culture and recombinant expressed

CC Mammary epithelial cell growth inhibitor in culture and recombinant expressed

CC useful as a therapeutic agent in the treatment of breast cancer

Sequence 538 AA;

Alignment Scores:

Pred. No.: 1.03e-76 Length: 538
Score: 1033.00 Matches: 212
Percent Similarity: 95.52% Conservative: 1
Best Local Similarity: 95.07% Mismatches: 7
Query Match: 24.42% Indels: 4
DB: 2 Gaps: 0

US-10-028-952a-4 (1-2333) x AA23756 (1-538)

```

QY      417 TTGATCTTGATTTTTCAGTACGAATACAGCCGGAAGGGGGCCCTCAGATCTCTGA 476
Db      1 MetIleLeuIlePheSerThrAsnTrpAspArgLysSerGlyAlaSer Asp LeuLeu 20
QY      477 CCTTTGGGTTTAAACAGAGAGGTGTCAAGAAAGTTTACACAGAGATTAATGCTTGTG 536
Db      20 hrPheTrpValLeuSerArgArgCysGlyHisSerTyHisArgAspAsnTrpLeuVal 40
QY      537 CGGCGAAGCTTATACGACACCGCGCTTTTATTCCTTGATGCTGCGCTTCTTATCAT 596
Db      40 IAlaIysArgSerLysArgArgPheLeuIleLeuArgCysArgLeuPheLeuSer 60
QY      597 TGTGAAGCAGAAATTCACCAAGCGTTGATTTTACCCCTAATAATAGGAACGTGACTG 656
Db      60 eugIyserArgIleHisGlnAlaLeuAspCysSerProThrAsnArgGluArgGluLeu 80
QY      657 GATTAGACCGCTCGTGAAGAGGTAGTTTACCCCTAATAATAGGAACGTGACTG 716
Db      80 ILeuAspArgArgGluTrpGly-LeuPheThrLeuLeuMetCysCysGlySerGly 99
QY      717 AATCTGCTCAGTACGAGAGAAACCGAGGTTGACATTTGTGTATGCTTGGCTGA 776
Db      100 AsnProAlaGlnTrpGluArgAsnArgArgPheArgHisLeuValTrpValLeuGlyTrp 119
QY      777 GAGCCCAATGAGGAGAACTTACCATCTGTGGATTTATGACTGAACGCTTATAGTGA 836
Db      120 GlyAlaAsnGlyAlaIysLeuProSerValGlyLeuLeuLeuAsnAlaSerLysSerGly 139
QY      837 TCCGCGCCAGAGGAGAACTTACCATCTGTGGATTTATGACTGAACGCTTATAGTGA 896
Db      140 SerArgProGlyGlyThrIleArgGlnArgArgGlyAlaSerValGlyLeuGlyTrpPro 159
QY      897 GTCCCGCCGCTGTCCCGCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955
Db      160 ValProArgLeuSerProProAlaGlyAlaProProProProArgAlaProAlaArg 179
QY      956 GAGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
Db      180 GluGlyAlaCysProAlaAlaArgArgAspArgGlyProValArgSerAlaLeuArgPro 199
QY      1016 GGGAAACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
Db      200 GlyLysArgGlyAlaAlaGlyLysAlaAlaAlaProSerProValThrHisArgThrPhe 219
QY      1076 GTG 1078
Db      220 Val 220

```

RESULT 6

AB804724 standard; protein; 302 AA.

AB804724;

11-MAR-2002 (first entry)

Human PPS423 protein SEQ ID NO:26.

Human; PPS423; cancer suppression.

Homo sapiens.

CN131316-A.

19-SEP-2001.

13-MAR-2000; 2000CN-00111990.

13-MAR-2000; 2000CN-00111990.

(SHAN-) SHANGHAI INST ONCOLOGY.

XX

PI Gu J, Yang S;
 XX WPI, 2002-042194/06.
 DR N-PSDB; ABA04459.
 XX
 PT New human protein able to suppress growth of cancer cells and its
 PT encoding polynucleotide.
 XX
 PS Claim 1; Page 35 (Disclosure); 38pp; Chinese.
 CC The present sequence represents human PP5423 protein, which has cancer
 CC suppressing activity. The present invention describes a method for the
 CC preparation of the protein by recombination, and the application of the
 CC protein in treating diseases such as cancer
 XX
 SQ Sequence 302 AA:
 Alignment Scores:
 Pred. No.: 1,98e-63 Length: 302
 Score: 872.00 Matches: 179
 Percent Similarity: 74.21% Conservative: 8
 Best Local Similarity: 71.03% Mismatches: 57
 Query Match: 20.61% Indels: 9
 DB: 5 Gaps: 2
 US-10-028-952a-4 (1-2333) X ABB04724 (1-302)
 QY 841 GCCAGCGGAGGAGATACGCGAGCGCGGAGCTCGGTTGGCTCGGATAGCCGATCC 900
 DB 53 AAlLysAlaLeuLys-----GluGluGluAsnArgArgLeuGluLysArgArgAla 70
 QY 901 CCGCGCTGTCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 960
 DB 71 GluGluAlaGlnArgGlnLeuGlnGlnArgArgGlnLeuArgGlnAlaAla 90
 QY 961 CGGTCGGAA 1020
 DB 91 ArgArgGluGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 110
 QY 1021 ACCGG-----GCGCGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCA 1061
 DB 111 GlnGlnGlnGlnValValSerArgAsnArgAsnGlnGlnGlnGlnGlnGlnGln 130
 QY 1062 CGGACCGCGAGCTCGTGGGAGACCTGGCGCTAAGCCTCCATCCGCTCGCTCGCT 1121
 DB 131 GluIlePheLysGlnLysGlnArgAlaMetSerThrThrSerThrSerThrSerThr 150
 QY 1122 GGCAGGTGAGAGACCCCTTCTGCAAGGCCCGAGATCTCCCTGCTGAGAGCCGGG 1181
 DB 151 GlnLysLeuArgSerProPheLeuGlnLysGlnLeuThrGlnProGlnThrHisPheGly 170
 QY 1182 AGAGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241
 DB 171 ArgGlnProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 190
 QY 1242 CCAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
 DB 191 ProSerThrProProCysLeuValGlnAlaGlnGlnAlaValAlaLysGlnLysPro 210
 QY 1302 GAGCAGGAGACCTTCTTACGAGCAGCCCGCCTGCTGCTGCTGCTGCTGCTGCTG 1361
 DB 211 GlnGlnGlnThrPheArgGlnGlnProProLeuValGlnGlnGlnGlnGlnGln 230
 QY 1362 CACATTGACACCACTTACGAGCAGCCCGCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1421
 DB 231 HisIleAspHisHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 250
 QY 1422 CTGTAGACTTACGAGCAGCCCGCAGACAGAGATCTCTTTGACCCCGAAGCTCATC 1481
 DB 251 LeuTyrAspTyrGlnAlaAlaAspAspThrGlnIleSerPheAspProGlnLysLeu 270
 QY 1482 ACCGGGATGAGGTGATCGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1541

DB 271 ThGlyIleGluValIleAspGlnGlyTrpTrpArgGlyTyrGlyProAspGlyIlePhe 290
 QY 1542 GGCATGTTCCCTGCCCAACTACGTGAGCTCATTTAG 1577
 DB 291 GlyMetPheProAlaAsnTyrValGluLeuIleGlu 302
 RESULT 7
 AAM78585
 ID AAM78585 standard; protein: 430 AA.
 XX
 AC AAM78585;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1247.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 OS nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 RN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HSE-) HXSEQ INC.
 XX
 PI Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QH, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI, 2001-476283/51.
 DR N-PSDB; AAK51718.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 20; Page 3507-3508; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK5581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 430 AA:
 Alignment Scores:
 Pred. No.: 2.18e-63 Length: 430
 Score: 872.00 Matches: 179
 Percent Similarity: 74.21% Conservative: 8
 Best Local Similarity: 71.03% Mismatches: 57

Query Match: 20.61% Indels: 9
 DB: 4 Gaps: 2
 US-10-028-952a-4 (1-2333) x AAM78585 (1-430)

QY 841 GCCCAGCGGGAACGATACGCGAGCCCGCGAGCCTCGGTTGGCTCGATAGCCGGTCC 900
 DB 181 AAlATySAlAGluTyS-----GluGluGluTubAmAArgLeGluGluTySArgAAla 198
 QY 901 CCCGCTGTCCCGCGCGGCGCGCCCGCTCCACCGCGCGCGCGCGCGGAGGG 960
 DB 199 GluGluAlaGlnArgGlnLeuGlnGlnGlnArgAArgGluLeuAlaAla 218
 QY 961 CGCGTGGCCCGGA 1020
 DB 219 ArgArgGluGlnArgTyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 238
 QY 1021 ACGGG-----CGCGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGTCA 1061
 DB 239 GlnGlnGlnGlnValValSerArgAmAArgAnglnGlnGlnSerAlaValHisProArg 258
 QY 1062 CGCAGCGCGCGTTCGCGGGAACCTGGCGCTAAACCATCTCCATCTCCAGCTT 1121
 DB 259 GluIlePheIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 278
 QY 1122 GCGAAGCTGAGAGCGCCCTCTCTGCGAAGAGCAGTCAACCAACGAGAGCCACTTGGC 1181
 DB 279 GlyIlePheIleArgSerProPheLeuGlnIleGlnIleGlnIleGlnIleGlnIle 298
 QY 1182 AGAGAGCGAGCTGTGCGCATCTTCAGAGCCAGGCGAGATCTCTCTGAGAGCGCGG 1241
 DB 299 ArgGluProAlaAlaAlaIleSerArgProArgAlaAspLeuProAlaGlnGluProAla 318
 QY 1242 CCAGAGCTCTCTCATGTCTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1301
 DB 319 ProSerThrProProCysLeuValGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGln 338
 QY 1302 GAGCAGAGAGAGCTTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1361
 DB 339 GluGlnGlnIleThrPheTyGlnGlnIleProLeuValGlnGlnGlnGlnGlnGlnGln 358
 QY 1362 CACATGACACACACATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1421
 DB 359 HisIlePhePheHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 378
 QY 1422 CTGTACGACTACGAGAGAGAGAGAGAGAGAGATCTCTTACCGCGAGAGCTCATC 1481
 DB 379 LeuTyAspTyGlnAlaAlaAlaAspThrGlnIleSerPheAspProGluAsnLeuIle 398
 QY 1482 ACCGCGCATCGAGGTATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1541
 DB 399 ThrGlyIleGlnValIleAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 418
 QY 1542 GGCATGTTCCCTGCGCACTACGTGAGCTCATTTGAG 1577
 DB 419 GlyMetPheProAlaAsnTyValGluLeuIleGlu 430

RESULT 8
 AAM79569 ID AAM79569 standard; protein; 458 AA.
 XX AAM79569;
 XX
 DT 06-NOV-2001 (first entry)
 XX Human protein SEQ ID NO 3215.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 XX nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.

XX
 PN W0200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WC-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR N-PSDB; AAK52702.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 20; Page 287; 6221pp; English.
 XX
 SS The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 458 AA;
 Alignment Scores:
 Pred. No.: 2.22e-63 Length: 458
 Score: 872.00 Matches: 179
 Percent Similarity: 74.21% Conservative: 8
 Best Local Similarity: 71.03% Mismatches: 57
 Query Match: 20.61% Indels: 9
 DB: 4 Gaps: 2

US-10-028-952a-4 (1-2333) x AAM79569 (1-458)

QY 841 GCCCAGCGGGAACGATACGCGAGCCCGCGAGCCTCGGTTGGCTCGATAGCCGGTCC 900
 DB 209 AAlATySAlAGluTyS-----GluGluGluTubAmAArgLeuGluGluTySArgAAla 226
 QY 901 CCCGCTGTCCCGCGCGGCGCGCCCGCTCCACCGCGCGCGCGCGCGGAGGG 960
 DB 227 GluGluAlaGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 246
 QY 961 CGCGTGGCCCGCGCGCGCGCGAGCGAGTCCGCTCGATAGCCGGTCCGCGGA 1020
 DB 247 ArgArgGluGlnArgTyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 266
 QY 1021 ACGGG-----CGCGGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGTCA 1061
 DB 267 GlnGlnGlnGlnValValSerArgAmAArgAnglnGlnGlnSerAlaValHisProArg 286
 QY 1062 CGCAGCGCGAGTTCGCGGGAACCTGGCGCTAAACCATCTCCATCTCCAGCTT 1121

```

Db      287  GluIlePheIysGlnIysGlnAArgAlaMetSerThrThrsSerIleSerSerProGlnPro 306
QY      1122  GGCAGAGCTGAGAGCCCTCTCTGCGAGAGACAGCTCAACCAACAGAGACCCACTTTGGC 1181
Db      307  GlyIysLeuArgSerProPheLeuGlnIysGlnLeuThrGlnProGlnThrHisPheGly 326
QY      1182  AGAGAGCAGCTGCTGCGATCTCAAGGCCAGGAGATCTCCCTGCTGAGAGCCGGCG 1241
Db      327  ArgGlnProAlaAlaAlaIleSerHisGlnProArgAlaIlePheProAlaGlnProAla 346
QY      1242  CCCAGCACTCTCCATCTGTGTGTCAGGAGCAAGAGAGAGGCTGTGTATGAGAACTCCA 1301
Db      347  ProSerThrProProCysLeuValGlnAlaGlnGlnGlnValAlaValTyrGlnGlnProPro 366
QY      1302  GAGAGAGAGACCTTCTTCAAGAGAGCCCACTGGTGTAGAGAGAGAGTGTGCTGTAG 1361
Db      367  GlnGlnGlnThrPheTyrGlnGlnProProLeuValGlnGlnGlnGlnValArgIysSerGln 386
QY      1362  CACATTGACCAACCAATTCAGGGCCAGGAGGCTCAATGTGAGCAAGGCTCTGTGCTGCTGC 1421
Db      387  HisIleAspHisHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 406
QY      1422  CTGAGACATCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1481
Db      407  LeuTyrAspTyrGlnAlaAlaAspAspThrGlnIleSerPheAspProGlnIleLeuIle 426
QY      1482  ACGGAGCATCGAGTGTATGAGAGAGGCTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1541
Db      427  ThrGlyIleGlnValIleAspGlnGlyTyrTyrPheGlyTyrGlyProAspGlyHisPhe 446
QY      1542  GGCATGTTCCCTGCGCACTACGTGAGAGCTCAATGAG 1577
Db      447  GlyMetPheProAlaAsnTyrValGlnLeuIleGln 458

RESULT 9
AAB63214
ID      AAB63214 standard; protein, 199 AA.
AC      AAB63214;
CC
XX
XX      26-MAR-2001 (first entry)
DT
XX
XX      Human secreted protein sequence encoded by gene 38 SEQ ID NO:140.
DE
XX
XX      Human, secreted protein; diagnosis; immunosuppressive; antiarthritic;
KM      antirheumatic; antiproliferative; cyostatic; cardiac; vasotropic;
KM      cerebroprotective; neuroprotective; antibacterial; virucide;
KM      fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm;
KM      rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KM      cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KM      angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KM      ocular disorder; corneal infection; wound healing; skin aging;
KM      food additive; preservative.
XX
XX      Homo sapiens.
OS
XX
XX      MO200061629-A1.
FN
XX
XX      19-OCT-2000.
PD
XX
XX      06-APR-2000; 2000MO-US009071.
PF
XX
XX      09-APR-1999; 99US-0128694P.
PR      20-JAN-2000; 2000US-0176931P.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
PA      (ROSE/) ROSEN C A.
XX
XX      Ruben SM, Komatsoulis G;
PI
XX
XX      MPI, 2000-647420/62.
XX

```

```

PT      Isolated nucleic acid molecule encoding a human secreted protein is used
PT      in preventing, treating or ameliorating a medical condition.
XX
XX      Disclosure; Page 518-519; 533pp; English.
PS
XX
XX      AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134
CC      to AAB63182. AAB63183 to AAB63231 represent more human secreted proteins
CC      and polypeptides homologous to them. Human secreted proteins have
CC      activities based on the tissues and cells the genes are expressed in.
CC      Examples of activities include: immunosuppressive; antiarthritic;
CC      antirheumatic; antiproliferative; cyostatic; cardiac; vasotropic;
CC      cerebroprotective; neuroprotective; antibacterial; virucide;
CC      fungicide; and ophthalmological. The polynucleotides and proteins can be
CC      used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC      mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
CC      also used in diagnosing a pathological condition or susceptibility to a
CC      pathological condition. Disorders which are diagnosed or treated include
CC      autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC      disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC      e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC      angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC      infections caused by bacteria, viruses and fungi and ocular disorders
CC      e.g. corneal infection. The polypeptides can also be used to aid wound
CC      healing and epithelial cell proliferation, to prevent skin aging due to
CC      sunburn, to maintain organs before transplantation, for supporting cell
CC      culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC      polypeptides can also be used as a food additive or preservative to
CC      increase or decrease storage capabilities. AAF22364 to AAF22372 and
CC      AAB63133 represent sequences used in the exemplification of the present
CC      invention
XX
SQ      Sequence 199 AA;
XX
XX
XX      Alignment Scores:
XX      Pred. No.: 2,586-63 Length: 199
XX      Score: 870.00 Matches: 161
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 20.57% Indels: 0
XX      DB: Gaps: 3 Gaps: 0

US-10-028-952A-4 (1-2333) x AAB63214 (1-199)
QY      1095  ACCACCTTCATCTCCAGTCTCTGAGCTGTGAGAGAGCCCTTCTCTGAGAGAGAG 1154
Db      39  ThrThrSerIleSerSerProGlnProGlyIysLeuArgSerProPheLeuGlnIysGln 58
QY      1155  CTCACCCCAACGAGAACCACTTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1214
Db      59  LeuThrGlnProGlnThrHisPheGlyArgGlnProAlaAlaAlaIleSerHisGlnProArg 78
QY      1215  GCAGATCTCCCTGCTGAGAGAGCCGAGCCAGCACTCTCTCAATGCTGTGTGAGGAGAGAA 1274
Db      79  AlaAspLeuProAlaGlnGlnProAlaProSerThrProProCysLeuValGlnIleGln 98
QY      1275  GAGAGAGCTGTGTATGAGAGAACTCCAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1334
Db      99  GlnGlnAlaValTyrGlnGlnProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 118
QY      1335  GTGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1394
Db      119  ValGlnGlnGlnGlnValAlaGlySerGlnHisIleAspHisHisIleGlnGlnGlnGlnGln 138
QY      1395  AGTGGGCAAGGAGCTGTGTGCTCCGAGGCTGTGAGCTCAAGAGAGGAGGAGGAGGAGGAGGAG 1454
Db      139  SerGlyGlnGlyLeuValAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspAspThrGln 158
QY      1455  ATCTCTTGTAGCCCGAGAACTCATACGAGGAGTGTGAGGAGTGTGAGGAGGAGGAGGAGGAG 1514
Db      159  IleSerPheAspProGlnLeuLeuIleThrGlyIleGlnValIleAspGlnGlnGlnGlnGln 178
QY      1515  CTGTGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1574

```

Db 179 ArgglyTyrGlyProaspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 198
 QY 1575 GAG 1577
 199 Glu 199
 Db
 RESULT 10
 ABB57455
 ID ABB57455 standard; protein; 203 AA.
 AC ABB57455;
 XX
 DT 15-MAR-2002 (first entry)
 DE Human secreted protein encoding polypeptide SEQ ID NO 101.
 XX
 XX Human; neurotrophic; cytoprotective; cytostatic; dermatological; vitruclide;
 KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatocrotic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antidiuretic; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
 XX
 OS Homo sapiens.
 XX
 XX WO200183510-A1.
 PN 08-NOV-2001.
 PD 26-APR-2001; 2001WO-US013318.
 XX
 PF 02-MAY-2000; 2000US-0201194P.
 PR 16-JUN-2000; 2000US-0212142P.
 PR
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Komatsoulis G, Ruben SM, Rosen CA;
 PI
 DR WPI; 2002-121886/16.
 XX
 PT An isolated nucleic acid molecule encoding a human secreted protein (SP)
 PT is useful in preventing, treating or ameliorating a disorder e.g.,
 PT Alzheimer's disease and cancers.
 XX
 PS Disclosure; Page 491; 496pp; English.
 XX
 XX The invention relates to novel genes (AB101566-AB101594) and proteins
 CC (ABB57394-ABB57456) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer; e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections
 CC
 XX
 XX Sequence 203 AA;
 SQ
 Alignment Scores:
 Pred. No.: 2.59e-63 Length: 203
 Score: 870.00 Matches: 161
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.57% Indels: 0
 DB: 5 Gaps: 0

US-10-028-952a-4 (1-2333) x ABB57455 (1-203)
 QY 1095 ACCACCTCCATCTCCAGTCCCTGAGCAAGCTGAGAGAGGCCCTTCTCGAGAGAGAG 1154
 Db 43 ThrThSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnGln 62
 QY 1155 CTCACCCAG 1214
 Db 63 LeuThrGlnProGlnThrHisPheGlyArgGlnProAlaAlaAlaIleSerArgProArg 82
 QY 1215 GCAGATCTCTCTGAG 1274
 Db 83 AlaAspLeuProAlaGlnGlnProAlaProSerThrProProGlyLeuValGlnGln 102
 QY 1275 GAGGAGGCTGTATGAG 1334
 Db 103 GluGlnAlaValTyrGlnGlnProProGlnGlnGlnThrPheTyrGlnGlnProProLeu 122
 QY 1335 GTCCAG 1394
 Db 123 ValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 142
 QY 1395 AGTGGGAG 1454
 Db 143 SerGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 162
 QY 1455 ATCTCTTTGACCCCGAG 1514
 Db 163 IleSerPheAspProGlnAsnLeuIleThrGlyIleGluValIleAspGlnGlyTyrPTrp 182
 QY 1515 CGTGGCTATGGGCGGAG 1574
 Db 183 ArgglyTyrGlyProaspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 202
 QY 1575 GAG 1577
 203 Glu 203
 Db
 RESULT 11
 ID AAY73941
 AAY73941 standard; protein; 217 AA.
 XX
 XX AAY73941;
 AC
 XX
 XX 14-MAR-2000 (first entry)
 DT
 DE Human prostate tumor EST fragment derived protein #128.
 XX
 XX Pancreas; tumor; EST; expressed sequence tag; human; cytoprotatic;
 KW treatment.
 XX
 XX Homo sapiens.
 OS
 XX DE19820190-A1.
 PN 04-NOV-1999.
 PD 28-APR-1998; 98DE-01020190.
 PF 28-APR-1998; 98DE-01020190.
 PR 28-APR-1998; 98DE-01020190.
 PR
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 XX
 XX WPI; 1999-621386/54.
 DR N-PSDB; AA252900.
 XX
 XX New human nucleic acid sequences from pancreatic tumors, and related
 PT proteins.
 XX
 XX Claim 23; Page 362; 502pp; German.
 XX

CC This invention describes novel polypeptides and their encoding nucleic acids derived from human pancreatic tumor tissue which have cytostatic activity. The sequences are also useful in producing pharmaceutical compositions for treatment of pancreatic tumors. AA73814-Y74252
 CC represent protein fragments encoded by the human pancreatic tumor cDNA library derived expressed sequence tag (EST) sequences represented in AA252858-Z53014

CC Sequence 217 AA;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	2,64e-63	217	161	0	0	0
Percent Similarity:	870.00					
Best Local Similarity:	100.00%					
Query Match:	20.57%					

US-10-028-952a-4 (1-2333) x AAV73941 (1-217)

QY 1095 ACCACCTCCATCTCCAGTCTCCAGCTGCGAGAGTGAAGAGCCCTTCTCGAGAGAGCAG 1154
 DB 57 ThrttserieteserProGlnProGlyLysleuArgserProPheleuGlnLysGln 76
 QY 1155 CTCACCCAG 1214
 DB 77 LeuthrGlnProGlnLysleuArgserProPheleuGlnLysGln 96
 QY 1215 GCAGATCTCCCTGCTGAG 1274
 DB 97 AlaAspLeuProGlnLysleuArgserProPheleuGlnLysGln 116
 QY 1275 GAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1334
 DB 117 GlnLysleuArgserProGlnLysleuArgserProPheleuGlnLysGln 136
 QY 1335 GTGAG 1394
 DB 137 ValInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 156
 QY 1395 AGTGGGAG 1454
 DB 157 SerLysleuArgserProGlnLysleuArgserProPheleuGlnLysGln 176
 QY 1455 ATTCCTTTCAG 1514
 DB 177 IleSerPheAspProGlnLysleuArgserProPheleuGlnLysGln 196
 QY 1515 CGTGGCTTATGGGCGGAGTGGCATTTTGGCATTTTGGCATTTTGGCATTTTGGCATTTT 1574
 DB 197 ArgGlyTyrGlyProAspGlyLysPheGlyMetPheProAlaAspTyrValGlnLeu 216
 QY 1575 GAG 1577
 DB 217 Gln 217
 RESULT 12
 ABP98850
 ID ABP98850 standard; protein; 377 AA.
 XX
 AC ABP98850;
 XX
 DT 15-JUN-2003 (first entry)
 XX
 DE Human structural and cytoskeletal associated protein #41.
 XX
 KW Cytostatic; antiarteriosclerotic; anticonvulsant; nocotropic; antianginal;
 KW neuroprotective; cerebroprotective; hypotensive; cardiact; osteoprotic;
 KW antiinflammatory; antiarthritic; virucide; gene therapy; human; stroke;
 KW structural and cytoskeleton-associated protein; SCAP; cancer; angina;
 KW arteriosclerosis; epilepsy; Huntington's disease; hypertension;
 KW heart failure; osteoporosis; osteoarthritis.

OS Homo sapiens.

XX Chavla NK, Elliott VS, Emerling BV, Forsythe II, Goryad AB, Lal PG;
 XX WO2003031940-A2.

XX 17-APR-2003.

XX 10-OCT-2002; 2002WO-US032851.

XX 12-OCT-2001; 2001US-0328931P.

XX 19-OCT-2001; 2001US-0360681P.

XX 02-NOV-2001; 2001US-0343896P.

XX 09-NOV-2001; 2001US-0343808P.

XX 16-NOV-2001; 2001US-0333385P.

XX 07-DEC-2001; 2001US-0340767P.

XX 11-JAN-2002; 2002US-0347703P.

XX (INCYTE GENOMICS INC.

XX Becha SD, Bhatia U, Blake JJ, Borowsky ML, Burrill JD, Chang H;
 XX Chavla NK, Elliott VS, Emerling BV, Forsythe II, Goryad AB, Lal PG;
 XX Griffin JA, Hafalia AD, Ho A, Ison CH, Kable AE, Katre R, Luo W;
 XX Lee S, Lee EA, Lee SY, Lehr-Mason PM, Li X, Lindquist EA, Luo W;
 XX Margulis JP, Ramkumar J, Richardson TW, Sprague WW, Swarnakar A;
 XX Tang YT, Warren BA, Yang J, Yue H, Zebardjian Y, Zheng W;
 XX WPI; 2003-403125/38.
 XX N-PSDB; ACC44337.

XX New human structural and cytoskeleton-associated proteins (SCAP) useful
 XX for diagnosing, treating and preventing diseases or conditions associated
 XX with the aberrant SCAP expression e.g. cancer, osteoporosis, or epilepsy.

XX Claim 1; Page 297; 361pp; English.

XX This sequence represents a novel isolated human structural and
 XX cytoskeleton-associated protein (SCAP) polypeptide. The polypeptides and
 XX polynucleotides encoding them are useful in diagnosing, treating and
 XX preventing diseases or conditions associated with the decreased
 XX expression or over expression of SCAP, such as cell proliferative (e.g.
 XX cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
 XX disease, stroke), heart (e.g. hypertension, heart failure, angina) and
 XX skeletal muscle disorders (e.g. osteoporosis, osteoarthritis) or viral
 XX infections. These are also useful in assessing the effects of exogenous
 XX compounds on the expression of nucleic acid and amino acid sequences of
 XX SCAP. The SCAP or its fragments are useful in screening compounds for
 XX effectiveness as agonist or antagonist of the polypeptides, or in
 XX altering the expression of the target polynucleotide and compounds that
 XX specifically bind to or modulate the activity of the polypeptide. The
 XX microarray is useful in monitoring or measuring protein-protein
 XX interactions, drug-target interactions, and gene expression profiles

XX Sequence 377 AA;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3,08e-63	377	161	0	0	0
Percent Similarity:	870.00					
Best Local Similarity:	100.00%					
Query Match:	20.57%					

US-10-028-952a-4 (1-2333) x ABP98850 (1-377)

QY 1095 ACCACCTCCATCTCCAGTCTCCAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1154
 DB 217 ThrttserieteserProGlnProGlyLysleuArgserProPheleuGlnLysGln 236
 QY 1155 CTCACCCAG 1214
 DB 237 LeuthrGlnProGlnLysleuArgserProPheleuGlnLysGln 256
 QY 1215 GCAGATCTCCCTGCTGAG 1274

Db 257 AlaAspLeuProAlaGluGluProAlaProSerThrProProCysLeuValGlnAlaGlu 276
 QY 1275 GAGAGAGCGCTGTGTATGAGAACTCCAGACGACGAGAACTTCTTACAGACAGCCCCCACTG 1334
 Db 277 GluGlnAlaValTyrGluGluProProGluGlnGluThrPheTyrGluGlnProProLeu 296
 QY 1335 GTGACGACGAGAGGTGTGCTGTGACACATTCAGACACCATTCAGAGGCGGAGCTC 1394
 Db 297 ValGlnGlnGlnGlnAlaGlySerGlnHisIleAspHisIleGlnGlnGlnGlyLeu 316
 QY 1395 AGTGGGCAAGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1454
 Db 317 SerGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 336
 QY 1455 ATCTCTCTTGAACCCCGAGAACTCATCAGCGGCGCATGAGGTGATGAGCGAGGCTGTGG 1514
 Db 337 IleSerPheAspProGluAsnLeuIleThrGlyIleGlnValIleAspGluGlyTyrTrp 356
 QY 1515 CGTGGCTATGAGCGCGAGATGCGCAATTTGGCATGTTCCTGCGCACTACGTGAGCTCAT 1574
 Db 357 ArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 376
 QY 1575 GAG 1577
 Db 377 Glu 377

RESULT 13

AA820896
 ID AAB20896 standard; protein; 431 AA.

XX AAB20896;

XX 08-JAN-2001 (first entry)

XX Human dreblin-like protein and SH3 domain sequence SEQ ID NO:1.

XX Human; dreblin-like protein; SH3 domain; cancer; diagnosis.

XX Homo sapiens.

XX JP2000197489-A.

XX 18-JUL-2000.

XX 07-JAN-1999; 99JP-00002254.

XX 07-JAN-1999; 99JP-00002254.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2000-545912/50.

XX N-PSDB; AAA92163; AAA92164.

XX New human protein having dreblin-like sequence and SH3 domain.

XX Claim 1; Page 5-6; 12p; Japanese.

XX The present sequence represents a human protein having a dreblin-like

XX sequence and SH3 domain. The protein and antibody against the protein

XX can be used in the diagnosis and the treatment of various diseases

XX including cancers

XX SQ Sequence 431 AA;

Alignment Scores:

Pred. No.: 3.2e-63 Length: 431
 Score: 870.00 Matches: 161
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.57% Indels: 0
 DB: 3 Gaps: 0

US-10-028-952a-4 (1-233) x AAB20896 (1-431)

QY 1095 ACCACCTCCATCTCCAGTCTCCAGCTGTGAGAGCTGAGAGCCCTTCTCCAGAGCAG 1154
 Db 271 ThrThrSerIleSerSerProGlnProGlyValLeuArgSerProPheLeuGlnValGln 290
 QY 1155 CTCACCCACACGAGAGACCCCATTTGGCAGAGAGGCACTGTGCTGCAATCTCAAGCCCAAG 1214
 Db 291 LeuThrGlnProGluThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProArg 310
 QY 1215 GCAATCTCTCTGCTGAGAGACCGGCGCCGCACTCTCCATGTCTGGTGCAGAGAA 1274
 Db 311 AlaAspLeuProAlaGluGluProAlaProSerThrProProCysLeuValGlnAlaGlu 330
 QY 1275 GAGAGAGCTGTGTATGAGAACTCCAGACGAGAGACCTTCTTACAGACAGCCCCCACTG 1334
 Db 331 GluGlnAlaValTyrGluGluProProGluGlnGluThrPheTyrGluGlnProProLeu 350
 QY 1335 GTGACGACGACGAGGTGCTGTGCTGTGACACATTCAGACACCATTCAGAGGCGGAGCTC 1394
 Db 351 ValGlnGlnGlnGlnAlaGlySerGlnHisIleAspHisIleGlnGlnGlnGlyLeu 370
 QY 1395 AGTGGGCAAGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1454
 Db 371 SerGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 390
 QY 1455 ATCTCTCTTGAACCCCGAGAACTCATCAGCGGCGCATGAGGTGATGAGCGAGGCTGTGG 1514
 Db 391 IleSerPheAspProGluAsnLeuIleThrGlyIleGlnValIleAspGluGlyTyrTrp 410
 QY 1515 CGTGGCTATGAGCGCGAGATGCGCAATTTGGCATGTTCCTGCGCACTACGTGAGCTCAT 1574
 Db 411 ArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 430
 QY 1575 GAG 1577
 Db 431 Glu 431

RESULT 14

AA85562
 ID AA85562 standard; protein; 431 AA.

XX AA85562;

XX 12-FEB-2001 (first entry)

XX Human tyrosine kinase substrate cks118/Dresh protein sequence.

XX Tyrosine kinase substrate; Tks; Tks 107, GRUB; Tks 113, Tks 118, Dresh;

XX Tks 202; Src; rheumatoid arthritis; artherosclerosis; stroke; cancer;

XX autoimmune disorder; organ transplantation; myocardial infarction;

XX cardiomyopathy; renal failure; neurodegenerative disorder;

XX cardiovascular disorder; inflammatory bowel disease; multiple sclerosis;

XX asthma; osteoarthritis; psoriasis; rhinitis.

XX Homo sapiens.

XX WO200061750-A2.

XX 19-OCT-2000.

XX 06-APR-2000; 2000MO-US009277.

XX 09-APR-1999; 99US-0128492P.

XX (SUG-) SUGEN INC.

XX Phan H, Courtneldge SA;

XX WPI; 2000-679486/66.

XX N-PSDB; AAC66078.

XX Novel tyrosine kinase substrate polypeptides and polynucleotides, used to

XX diagnose and treat diseases such as rheumatoid arthritis,

US-10-028-952a-4 (1-2333) x AAB93895 (1-439)

QY	1095	ACCACCTCATCTCCAGCTCTGAGGCAAGTGAAGAGCCCTTCTGTGAGAGCAG	1154
Db	279	ThrThrSerIleSerSerProGlnProGlyIleLeuArgSerProPheLeuGlnIleGln	298
QY	1155	CTGACCCCAACGAGAGCCCACTTTGGCAGAGAGCCAGCTGCTCCATCTCAAGGCCCAG	1214
Db	299	LeuThrGlnProGlnThrHisPheGlyArgGluProAlaAlaIleSerArgProArg	318
QY	1215	GCAGATCTCCCTGTGAGAGCGCGCCAGCACTCTCATGTCTGGTGCAGGCAGAA	1274
Db	319	AlaAspLeuProAlaGlnIleProAlaProSerThrProProCysLeuValGlnIleGln	338
QY	1275	GAGAGGCTGTGTATAGAGAACTCCAGAGCAGAGAGACTTCTAGAGCAGGCCCACTG	1334
Db	339	GlnGlnAlaValIleGlnIleProGlnIleProGlnIleThrPheIleGlnIleProProLeu	358
QY	1335	GTGCAAGCAGAGAGGCTGTGCTGTGAGCAATTGACCACTTCAAGGCGCCAGGGGCTC	1394
Db	359	ValGlnGlnIleGlnIleValIleGlnIleHisIleAspHisIleIleGlnIleGlnIleLeu	378
QY	1395	AGTGGCAAGGCTGTGTGCTGTGCTGTGAGCACTTACAGGAGCCGAGCAGACAGAG	1454
Db	379	SerGlyGlnGlyLeuGlnCysAlaArgAlaLeuIleAspIleGlnAlaAlaAspAspThrGln	398
QY	1455	ATCTCTTTGACCCCGAGAACTCATCAGGCGCATGAGGTATGACGAGAGGCTGTGG	1514
Db	399	IleSerPheAspProGlnLeuIleThrGlyIleGlnValIleAspGlnGlyIlePhe	418
QY	1515	CGTGGCTATGGCCGAGTGGCAATTTGGCATGTTCCTGCCAATACGTGAAGCTCATT	1574
Db	419	ArgGlyIleGlyProAspGlyHisPheGlyMetPheProAlaAsnIleValGlnLeuIle	438
QY	1575	GAG 1577	
Db	439	Glu 439	

Search completed: April 1, 2004, 15:53:09
Job time : 130.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:23:34 ; Search time 53.8053 Seconds

(without alignments)
4164.275 Million cell updates/sec

Title: US-10-028-952a-9

Perfect score: 3989

Sequence: 1 HEIPVPIVTPYPAKQXKXERA.....DFKSKKKKKLKLMTSSX 793

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: A_Geneseq29Jtn04:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3535	88.6	760	4 AAB66394	Aab66394 Human pro
2	1193.5	29.9	753	4 AAB66395	Aab66395 Human mam
3	800.5	20.1	355	2 AAM05398	Aam05398 Human clo
4	796	20.0	199	3 AAB63214	Aab63214 Human sec
5	796	20.0	203	5 ABB57455	Abb57455 Human sec
6	796	20.0	217	2 AAY73941	Aay73941 Human pro
7	796	20.0	302	5 ABB04724	Abb04724 Human pps
8	796	20.0	430	4 AAM78585	Aam78585 Human pro
9	796	20.0	431	3 AAB20896	Aab20896 Human dre
10	796	20.0	431	3 AAY85662	Aay85662 Human tyr
11	796	20.0	458	4 AAM79569	Aam79569 Human pro
12	794	19.9	377	6 AAB98850	Aab98850 Human str
13	794	19.9	439	4 AAB93895	Aab93895 Human pro
14	746.5	18.7	525	6 ABB52367	Abb52367 Human gpc
15	647	16.2	162	2 AAY73942	Aay73942 Human pro
16	628.5	15.8	433	2 AAM05389	Aam05389 Mouse SH3
17	626.5	15.7	201	3 AAB63213	Aab63213 Gene 38 h
18	409	10.3	538	2 AAY23756	Aay23756 Human mam
19	300	7.5	358	2 AAM34223	Aam34223 SH3 doma
20	300	7.5	58	2 AAM34248	Aam34248 SH3 doma
21	241	6.0	55	3 AAB53794	Aab53794 Human col
22	239.5	6.0	301	4 AAM83949	Aam83949 Human imm
23	234	5.9	486	6 ABR44110	Abr44110 Human pro
24	234	5.9	486	6 ABO07214	Abo07214 Human p53
25	234	5.9	486	7 ADD67624	Add67624 Human ly1

26	231	5.8	559	4 ABB60041	Abb60041 Drosophil
27	228.5	5.7	486	2 AAY03795	Aay03795 Murine lc
28	227.5	5.7	531	4 ABB71831	Abb71831 Drosophil
29	221.5	5.6	486	2 AAM05408	Aam05408 Mouse HSI
30	209	5.2	471	7 ABB55094	Abb55094 Rat Prote
31	208	5.2	546	6 AAM05406	Aam05406 Mouse cor
32	208	5.2	546	6 AAE38430	Aae38430 Mouse cor
33	205	5.1	456	6 AAE38429	Aae38429 Human cor
34	205	5.1	513	7 ADE55096	Ade55096 Human pro
35	202	5.1	550	6 ABO07213	Abo07213 Human p53
36	192	4.8	40	3 ABB53737	Abb53737 Human col
37	180.5	4.5	53	2 AAY03828	Aay03828 LckBPI al
38	177.5	4.4	53	6 AAE38444	Aae38444 Human cor
39	177.5	4.4	53	6 AAE38445	Aae38445 Mouse cor
40	176.5	4.4	54	2 AAY03827	Aay03827 LckBPI al
41	175.5	4.4	54	2 AAY03829	Aay03829 LckBPI al
42	171.5	4.3	53	2 AAY03821	Aay03821 LckBPI al
43	171.5	4.3	53	2 AAY03826	Aay03826 LckBPI al
44	171.5	4.3	53	2 AAY03820	Aay03820 LckBPI SH
45	170.5	4.3	53	2 AAY03824	Aay03824 LckBPI al

ALIGNMENTS

RESULT 1
AAB66394 standard; protein; 760 AA.
XX
XX AAB66394;
XX
DT 10-APR-2001 (first entry)
XX
DE Human prostate ECGI protein sequence.
XX
KW Human; mammaplastin A; mammaplastin B; mammaplastin C; ECGI; cancer;
XX epithelial cell growth inhibitor.
XX
OS Homo sapiens.
XX
XX WO200078955-A1.
XX
XX PD 28-DEC-2000.
XX
XX PF 19-JUN-2000; 2000WO-US016900.
XX
XX PR 16-JUN-1999; 99US-0139995P.
XX
XX PA (BIOT-) BIOTHERAPIES INC.
XX
XX PI Erlvin PR;
XX
XX DR WPI, 2001-071393/08.
XX N-PSDB; AAF31282.
XX
XX A Mammaplastin-like epithelial cell growth inhibitor, and the nucleic that
XX PT encodes it, useful for diagnosing and/or preventing epithelial cell
XX PT cancers, e.g. of the ovaries or prostate.
XX
XX PS Disclosure; Page 41-43; 55pp; English.
XX
XX CC The present invention describes a mammaplastin-like epithelial cell growth
XX CC inhibitor (ECGI) which has substantial similarity to mammaplastin A, B or
XX CC C. ECGI is expressed in healthy cells, but is either absent or expressed
XX CC at reduced levels in cancerous cells. The protein and its coding sequence
XX CC can be used to inhibit epithelial cell growth and the amount present in
XX CC cells can be used to diagnose cancer or monitor its progress
XX
XX SQ Sequence 760 AA;
XX
XX Query Match 88.6%; Score 3535; DB 4; Length 760;
XX Best Local Similarity 95.8%; Pred. No. 0;
XX Matches 759; Conservative 0; Mismatches 1; Indels 32; Gaps 30;

QY 1 HEIPIVPTIYPAKPOXKERAMNQGKTLSTLWHEEETXYXNKKYAFQAPV9PRG 60
 DB 1 HEIPIVPTIYPAKPOXKERAMNQGKTLSTLWHEEETXYXNKKYAFQAPV9PRG 60
 QY 61 ARGGXRPCCGPPPKYHYSDRFTDPRACGEPGALAXAKRPAARRPGATRGXXLRWG 120
 DB 61 ARGGXRPCCGPPPKYHYSDRFTDPRACGEPGALAXAKRPAARRPGATRGXXLRWG 117
 QY 121 VXLGRTTCOTVXQVSGELEDNPMKXKAPARLILFSTNTDXXSAGSEFPFGXA 180
 DB 121 VXLGRTTCOTVXQVSGELEDNPMKXKAPARLILFSTNTDXXSAGSEFPFGXA 175
 QY 181 GVKRKYTTGITGMRSVSDVAFXSPDVSSYHXXAEFTKRMIVHPLIGXSMDXVVR 240
 DB 176 G-VKATVTGITGMRSVSDVAFXSPDVSSYH-XAEFTKRMIVHPLIG-SMDXIVVR 232
 QY 241 QVSFTLLMKCCCHGNPAQYERNRXXHLVTVGXANGAKLSVGLXINAKSERSPGXT 300
 DB 233 QVSFTLLMKCCCHGNPAQYERNR-XHLVTVGXANGAKL-SVGLXINAKSERSPGXT 290
 QY 301 IRORRGAASYGLGXEXXRLSPAPRPPLNAPRGRARAPRAGTGVRCXVFPVLGNGAP 360
 DB 291 IRORRGAASYGLGX-XRLSPAPRPPLNAPR-RGRARAPRAGTGVRCXVFPVLGNGAP 348
 QY 361 ERFPKXRPRTARASWGTWRTXYSISPOPGKLSFFLOXOLTQPTPHGREPAAKXRP 420
 DB 349 ERFP-XRPRTARASWGTW-TSISPOPGKLSFFLOXOLTQPTPHGREPAA-XSRPR 404
 QY 421 ADLPAEPAPSPPCIVQAEBAVYEEPEXCEFTPEYQCPVLVQOQXGSEHIDH1QOQGL 480
 DB 405 ADLPAEPAPSP-PCIVQAEBAVYEEPEXCEFTPEYQCPVLVQOQ-XSEHIDH1QOQGL 462
 QY 481 SMOGLCARLIDYQADDDXISFDPENLITGIEVXXEGWAGYGPDDHFGKXPAHYELI 540
 DB 463 S-OGLCARLIDYQADDDXISFDPENLITGIEV-XEGWAGYGPDDHFGK-PAHYELI 519
 QY 541 EXGXGHIILFPQOTWLPYCMKRRPXXXSALFOEXDPOXGXKLRAPSGLADSAHPKCN 600
 DB 520 EXGXGHIILFPQOTWLPYCMKRRP-XXXSALFOEXDPOXG-XLAPSGLADSAHPKCN 577
 QY 601 GLVTPHPCIPRPPXXTAMLPLTGYXAKPCWBPSPPLPSCEGSGXGASGLMLPSA 660
 DB 578 GLVTPHPCIPR--TAMLPLTGYXAKPC-WBPSPPLPSCEGSGXGASGLMLPSA 634
 QY 661 FICLXXFSLASGWWPPLFRMNLGNSERRELFLEFVTVXVVDHGLAAGNLSCXHLCA 720
 DB 635 FICL-XFSLASGWWPPLFRM-LGNSERRELFLEFVTVXVVDHGLAAGNLSC-XHLCA 691
 QY 721 PHSISLSTGLGXGKGCRRWPSHPRYSKADTTGSGTFLTRCLQXXYTCASITDSIFRSGK 780
 DB 692 PHSISLSTGLG-GKGCRRWPSHPRYSKADTTGSGTFLTRCLQ-XVCASITDSIFRSGK 749
 QY 781 KXKLEKLTSS 792
 DB 750 K-KLEKLTSS 760

RESULT 2
 AAB66395
 ID AAB66395 standard; protein; 753 AA.
 AC AAB66395;
 DT 10-APR-2001 (first entry)
 DE Human mammatstatin C protein sequence.
 KW Human; mammatstatin A; mammatstatin B; mammatstatin C; ECGI; cancer;
 OS epithelial cell growth inhibitor.
 XX Homo sapiens.
 XX

PN W0200078955-A1.
 XX 28-DEC-2000.
 PD 19-JUN-2000; 2000WO-US016900.
 PF 18-JUN-1999; 99US-0139995P.
 PR (BIOT-) BIOTHERAPIES INC.
 PA Evin PR;
 PI WPI; 2001-071393/08.
 DR
 XX A Mammatstatin-like epithelial cell growth inhibitor, and the nucleic acid
 PT encodes it, useful for diagnosing and/or preventing epithelial cell
 PT cancer, e.g. of the ovaries or prostate.
 PS Disclosure, Page 44-46; 55pp; English.
 XX
 CC The present invention describes a mammatstatin-like epithelial cell growth
 CC inhibitor (ECGI) which has substantial similarity to mammatstatin A, B or
 CC C. ECGI is expressed in healthy cells, but is either absent or expressed
 CC at reduced levels in cancerous cells. The protein and its coding sequence
 CC can be used to inhibit epithelial cell growth and the amount present in
 CC cells can be used to diagnose cancer or monitor its progress
 XX
 SQ Sequence 753 AA;
 Query Match 29.9%; Score 1193.5; DB 4; Length 753;
 Best Local Similarity 81.6%; Pred. No. 1.6e-113;
 Matches 280; Conservative 1; Mismatches 29; Indels 33; Gaps 13;
 QY 37 HGEETXZYXNWXAPGAPPVSPGARGGXRPCCGPPPKYHYSDRFTDPRACGEPGAL 96
 DB 5 HGEETXZYXNWX-APGAPPVSPGARGGXRPCCGPPPKYHYSDR-XTDPRACGEPGAL 62
 QY 97 ASXAKRPAARRPGATRGSGXARWGVXLRGYTCQTVXQVSGELEDNPMKXKAPARLIL 156
 DB 63 AG-ARPAARRPGATRGSGXARWGVXLRGYTCQ-QVSGELEDNPMK-RRAPARLIL 119
 QY 157 LIFSTNTDXXSAGSARSTYPPGFYXGKRYKTTGTTGMRSVSDVAFXSPDVSSYHXX 216
 DB 120 LIFSTNTDXXSAGSARSTYPPGFYX-XRYKTTGTTGMRPS-HSDVAFXSPDVSSYHXX 177
 QY 217 AEFTRKMTVHPLIGXKXMDXTVVRQVSFTLLMKCCCHGNPAQYERNRXXHLVTVLGXGA 276
 DB 178 AEFTRKMTVHPLIG-XKXMDXTVVRQVSFTLLM-CCCHGNPAQYERNRXXHLVTVLGXGA 235
 QY 277 NGAKLXSVGLXINAKSERSPGKTIIRORRGAASYGLGXPPXRLSPAPRPPLNAPRGRGA 336
 DB 236 NGAR-XSVGLXINAKSERSPG-TIIRORRGAASYGLGXPPXRLSPAPRPPLNAPRGRGA 290
 QY 337 RAPPAGTGVRCXVFPVLGNGARPERRPXRPRTARASWGTW 379
 DB 291 -----GGRVPRAPRPGSACPSMSWETGR 314

RESULT 3
 AAM05398
 ID AAM05398 standard; protein; 355 AA.
 AC AAM05398;
 DT 19-FEB-1998 (first entry)
 DE Human clone 56 protein.
 KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
 KW cellular signaling element; cellular structural element; malignancy;
 KW protein identification; functional domain; protein screening;
 XX cellular signal transduction process.
 XX

OS Homo sapiens.
 XX WO9631625-A1.
 PN 10-OCT-1996.
 PD
 XX
 PF 04-APR-1996; 96WO-US004454.
 XX
 XX 07-APR-1995; 95US-00417872.
 PR 03-APR-1996; 96US-00630915.
 XX
 XX (CYTO-) CYTOGEN CORP.
 PA (UNIC-) UNIV NORTH CAROLINA.
 PI Sparks AB, Hoffman N, Kay BK, Fowlkes DW, McConnell SJ,
 XX WPI; 1996-465045/46.
 DR N-PSDB; AAT39798.
 XX
 PT Identifying polypeptide(s) having specific functional domain (esp. SH3
 PT domain) - comprises detecting selective binding to recognition unit,
 PT regardless of sequence homology.
 XX
 XX Claim 54; Fig 57; 174pp; English.
 PS
 XX AAW05386-W05403 represent novel human and mouse Src-homology region 3
 CC (SH3) domain containing proteins that can be used in the method of the
 CC invention. SH3 domain containing proteins play a role in signalling and
 CC structural elements of cells. The method of the invention is for
 CC identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). The method comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins which contain an SH3 domain due to the minimal sequence homology
 CC among known SH3 proteins. It has been found that small peptide RUs in
 CC multivalent form have reduced specificity for a given functional domain
 CC compared to monomer RUs. Multivalent RU complexes are particularly suited
 CC to screening for polypeptides containing functional domains that are
 CC similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention
 CC
 CC Sequence 355 AA;
 SQ
 Query Match 20.1%; Score 800.5; DB 2; Length 355;
 Best Local Similarity 64.9%; Pred. No. 1.9e-73;
 Matches 157; Conservative 9; Mismatches 47; Indels 29; Gaps 1;
 QY
 DB 359 RPERPXXRPRTRASNGTWTXSTSSPOGKRSRPTLOXLOLPETHREPREAAXSR 418
 112 RNEQSAVHPREIFQKRAMSTSTSSPOGKRSRPTLOXLOLPETHREPREAAXSR 171
 QY
 DB 419 PRADPAEPAPSPXPCVQAEAEVVEEPEYOEFTFEQPPPLVQOXXGSEHIDH1IQGQ 478
 172 PRADPAEPAPSPXPCVQAEAEVVEEPEYOEFTFEQPPPLVQOXXGSEHIDH1IQGQ 231
 QY
 DB 479 GLSXGGLCARALYDVQADDXEISPDENLITGIBYXKEGWWRVGPDGHGKMPAYVE 538
 232 GLSXGGLCARALYDVQADDXEISPDENLITGIBYXKEGWWRVGPDGHGKMPAYVE 291
 QY
 DB 539 LIEXKXGHIIXFP-----SQTWLPYCKRRPXXXXX 569
 292 LIEXKXGHIIXFP-----SQTWLPYCKRRPXXXXX 351
 QY
 DB 570 AL 571
 352 AM 353

RESULT 4
 AAB63214
 ID AAB63214 standard; protein; 199 AA.
 XX
 XX
 AC AAB63214;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 38 SEQ ID NO:140.
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytoskeletal; cardiac; vasotropic;
 KW cerebroprotective; neuroprotective; antibacterial; virocidic;
 KW fungicide; ophtalmological; gene therapy; autoimmune disease; neoplasm;
 KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
 KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
 KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
 KW ocular disorder; corneal infection; wound healing; skin aging;
 KW food additive; preservative.
 XX
 OS Homo sapiens.
 XX
 XX WO200061623-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US009071.
 XX
 PR 09-APR-1999; 95US-0128694P.
 XX
 PR 20-JAN-2000; 2000US-0176931P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PI Ruben SM, Komatseculis G;
 PT WPI; 2000-647420/62.
 PT
 PT Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 PS Disclosure; Page 518-519; 533pp; English.
 XX
 XX AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134
 CC to AAB63182. AAB63183 to AAB63231 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytoskeletal; cardiac; vasotropic;
 CC cerebroprotective; neuroprotective; antibacterial; virocidic;
 CC fungicide; and ophtalmological. The polynucleotides and proteins can be
 CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
 CC also used in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for chemotaxis. The
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities. AAF22364 to AAF22372 and
 CC AAB63133 represent sequences used in the exemplification of the present
 CC invention
 XX
 SQ Sequence 199 AA;

Query Match 20.0%; Score 796; DB 3; Length 199;
 Best Local Similarity 82.5%; Pred. No. 2.6e-73;
 Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPRXKPSRTASWGTWXTXSSISPOPKLSPLQXOLTQPTHTFGREPAAXSR 418
 17 RNEQSAVHPRRIFFQKERAMSTTSSISPOPKLSPLQXOLTQPTHTFGREPAAXSR 76
 QY 419 PRADLPAEPAPSPPCLVQAEAEVYEEPEQEFTFYEQPLVVOQXGSEHIDHIOGQ 478
 DB 77 PRADLPAEPAPSPPCLVQAEAEVYEEPEQEFTFYEQPLVVOQXGSEHIDHIOGQ 136
 QY 479 GLSXQGLCARALYDYOADDXEISPDPENLITGIEVXXEGWWRGYPDGHFGMPANYVE 538
 DB 137 GLSXQGLCARALYDYOADDXEISPDPENLITGIEVDEGWMWRGYPDGHFGMPANYVE 196
 QY 539 LIE 541
 DB 197 LIE 199

RESULT 5

AB57455
 ID AB57455 standard; protein; 203 AA.

AC AB57455;

DT 15-MAR-2002 (first entry)

DE Human secreted protein encoding polypeptide SEQ ID NO 101.

XX Human; noctropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vlnerary;
 KW antiparkinsonian; antischizling; antianaemic; antiahrtic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
 KW antiparastic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

OS Homo sapiens.

XX MO200183510-A1.

PD 08-NOV-2001.

PF 26-APR-2001; 2001MO-US013318.

PR 02-MAY-2000; 2000US-0201194P.

PR 16-JUN-2000; 2000US-0212142P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Komatsoulis G, Ruben SM, Rosen CA;

DR WPI; 2002-121886/16.

PT An isolated nucleic acid molecule encoding a human secreted protein (SP)
 PT is useful in preventing, treating or ameliorating a disorder e.g.,
 PT Alzheimer's disease and cancers.

PS Disclosure; Page 491; 496pp; English.

XX The invention relates to novel genes (AB101566-AB101594) and proteins
 CC (AB57394-AB57456) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)

CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
 CC (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections

XX Sequence 203 AA;

QY 359 RPERPRXKPSRTASWGTWXTXSSISPOPKLSPLQXOLTQPTHTFGREPAAXSR 418
 21 RNEQSAVHPRRIFFQKERAMSTTSSISPOPKLSPLQXOLTQPTHTFGREPAAXSR 80
 QY 419 PRADLPAEPAPSPPCLVQAEAEVYEEPEQEFTFYEQPLVVOQXGSEHIDHIOGQ 478
 DB 81 PRADLPAEPAPSPPCLVQAEAEVYEEPEQEFTFYEQPLVVOQXGSEHIDHIOGQ 140
 QY 479 GLSXQGLCARALYDYOADDXEISPDPENLITGIEVXXEGWWRGYPDGHFGMPANYVE 538
 DB 141 GLSXQGLCARALYDYOADDXEISPDPENLITGIEVDEGWMWRGYPDGHFGMPANYVE 200
 QY 539 LIE 541
 DB 201 LIE 203

RESULT 6

AA73941
 ID AAY73941 standard; protein; 217 AA.

XX AAY73941;

AC 14-MAR-2000 (first entry)

DT Human prostate tumor EST fragment derived protein #128.

XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
 KW treatment.

XX Homo sapiens.

OS DE19820190-A1.

PD 04-NOV-1999.

PF 28-APR-1998; 98DE-01020190.

PR 28-APR-1998; 98DE-01020190.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI; 1999-621386/54.

PT New human nucleic acid sequences from pancreatic tumors, and related
 PT proteins.

PS Claim 23; Page 362; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic
 CC acids derived from human pancreatic tumor tissue which have cytostatic
 CC activity. The sequences are also useful in producing pharmaceutical
 CC compositions for treatment of pancreatic tumors. AAY73814-174252
 CC represent protein fragments encoded by the human pancreatic tumor cDNA
 CC library derived expressed sequence tag (EST) sequences represented in
 CC AA52868-253014
 CC Sequence 217 AA;

Query Match 20.0%; Score 796; DB 2; Length 217;

Best Local Similarity 82.5%; Pred. No. 2.9e-73; Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPPXRRSRTKRSCTWXTXSSISPPGKRSFPLQXLTQPTTHGRRPAXXSR 418
DB 35 RNEQSAVHPRRIKOKERAMSTISISSPPGKRSFPLQXLTQPTTHGRRPAXXSR 94
QY 419 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEFTFEOPPLVQOQXGSEHIDHHIQG 478
DB 95 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEFTFEOPPLVQOQXGSEHIDHHIQG 154
QY 479 GLSXQGLCARALYDYQAADDXEISFPDENLITGIEVXXEGWMRGYGPDGHFQKXPANYVE 538
DB 155 GLSGQGLCARALYDYQAADDXEISFPDENLITGIEVDEGWMRGYGPDGHFQKXPANYVE 214
QY 539 LIE 541
DB 215 LIE 217

RESULT 7

ABB04724 ID ABB04724 standard; protein; 302 AA.

XX AC ABB04724;
XX DT 11-MAR-2002 (first entry)
XX XX Human PPS423 protein SEQ ID NO:26.
XX KW Human; PPS423; cancer suppression.
XX OS Homo sapiens.
XX PN CN1313316-A.
XX PD 19-SEP-2001.
XX PR 13-MAR-2000; 2000CN-00111990.
XX PR 13-MAR-2000; 2000CN-00111990.
XX PR 13-MAR-2000; 2000CN-00111990.
XX PA (SHAN-) SHANGHAI INST ONCOLOGY.
XX PI Gu J, Yang S;
XX DR WPI; 2002-042194/06.
XX DR N-PSDB; ABA04459.
XX PT New human protein able to suppress growth of cancer cells and its
XX PT encoding polynucleotide.
XX PS Claim 1; Page 35 (Disclosure); 38pp; Chinese.
XX CC The present sequence represents human PPS423 protein, which has cancer
XX CC suppressing activity. The present invention describes a method for the
XX CC preparation of the protein by recombinant, and the application of the
XX CC protein in treating diseases such as cancer
SQ Sequence 302 AA;

Query Match 20.0%; Score 796; DB 5; Length 302;

Best Local Similarity 82.5%; Pred. No. 4.5e-73; Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPPXRRSRTKRSCTWXTXSSISPPGKRSFPLQXLTQPTTHGRRPAXXSR 418
DB 120 RNEQSAVHPRRIKOKERAMSTISISSPPGKRSFPLQXLTQPTTHGRRPAXXSR 179
QY 419 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEFTFEOPPLVQOQXGSEHIDHHIQG 478
DB 180 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEFTFEOPPLVQOQXGSEHIDHHIQG 239

QY 479 GLSXQGLCARALYDYQAADDXEISFPDENLITGIEVXXEGWMRGYGPDGHFQKXPANYVE 538
DB 240 GLSGQGLCARALYDYQAADDXEISFPDENLITGIEVIDEGWMRGYGPDGHFQKXPANYVE 299

QY 539 LIE 541
DB 300 LIE 302

RESULT 8

AAM78585 ID AAM78585 standard; protein; 430 AA.

XX AC AAM78585;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein SEQ ID NO 1247.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US004098.

XX PR 03-FEB-2000; 2000US-00486914.
XX PR 27-APR-2000; 2000US-00560875.
XX PR 20-JUN-2000; 2000US-00598075.
XX PR 19-JUL-2000; 2000US-00620325.
XX PR 15-SEP-2000; 2000US-00654936.
XX PR 15-SEP-2000; 2000US-00653561.
XX PR 20-OCT-2000; 2000US-00683325.
XX PR 30-NOV-2000; 2000US-00728422.

XX PA (HYSB-) HYSBQ INC.

XX PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX PI Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX DR WPI; 2001-476283/51.
XX DR N-PSDB; AAK51718.

XX PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX PT in diagnosis and gene therapy.

XX PS Claim 20; Page 3507-3508; 6221pp; English.

XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activity/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
XX CC sequence listing were missing at the time of publication

SQ Sequence 430 AA;

Query Match 20.0%; Score 796; DB 4; Length 430;

Best Local Similarity 82.5%; Pred. No. 7.1e-73; Indels 0; Gaps 0;
Matches 151; Conservative 2; Mismatches 30;

QY 359 RPERRPXKRPRTARSMQWTWXTXSSPPGKLRSPFLQXLTQTPETHRGREPAAXSR 418
DB 248 RNEQBSAVHPREIFKQKRAMSTSSISPPGKLRSPFLQXLTQTPETHRGREPAAXSR 307

QY 419 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFYEQPLVVOQXGSEHIDH1QSQ 478
DB 308 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFYEQPLVVOQXGSEHIDH1QSQ 367

QY 479 GLSXQGLCARALYDQAADDEXISPDENLITGIEVXXEGMWRGYPDGHFGKXPANYVE 538
DB 368 GLSXQGLCARALYDQAADDEXISPDENLITGIEVXXEGMWRGYPDGHFGKXPANYVE 427

QY 539 LIE 541
DB 428 LIE 430

RESULT 9
AAB20896
ID AAB20896 standard; protein; 431 AA.
AC AAB20896;
XX
XX
DT 08-JAN-2001 (first entry)
DE Human dreblin-like protein and SH3 domain sequence SEQ ID NO:1.
XX
XX Human; dreblin-like protein; SH3 domain; cancer; diagnosis.
XX
XX Homo sapiens.
OS
XX JP2000197489-A.
XX
XX 18-JUL-2000.
XX
XX 07-JAN-1999; 99JP-00002254.
XX
XX 07-JAN-1999; 99JP-00002254.
XX
XX (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
XX
XX WPI; 2000-545912/50.
XX
XX DR N-PSDB; AAA92163, AAA92164.
XX
XX New human protein having dreblin-like sequence and SH3 domain.
XX
XX Claim 1; Page 5-6; 12pp; Japanese.
XX
XX The present sequence represents a human protein having a dreblin-like
CC sequence and SH3 domain. The protein and an antibody against the protein
CC can be used in the diagnosis and the treatment of various diseases
CC including cancers
CC
XX
XX Sequence 431 AA;

Query Match 20.0%; Score 796; DB 3; Length 431;
Best Local Similarity 82.5%; Pred. No. 7.1e-73; Indels 0; Gaps 0;
Matches 151; Conservative 2; Mismatches 30;

QY 359 RPERRPXKRPRTARSMQWTWXTXSSPPGKLRSPFLQXLTQTPETHRGREPAAXSR 418
DB 249 RNEQBSAVHPREIFKQKRAMSTSSISPPGKLRSPFLQXLTQTPETHRGREPAAXSR 308

QY 419 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFYEQPLVVOQXGSEHIDH1QSQ 478
DB 309 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFYEQPLVVOQXGSEHIDH1QSQ 368

QY 479 GLSXQGLCARALYDQAADDEXISPDENLITGIEVXXEGMWRGYPDGHFGKXPANYVE 538
DB 369 GLSXQGLCARALYDQAADDEXISPDENLITGIEVXXEGMWRGYPDGHFGKXPANYVE 428

QY 539 LIE 541
DB 429 LIE 431

RESULT 10
AA85662
ID AA85662 standard; protein; 431 AA.
AC AA85662;
XX
XX
DT 12-FEB-2001 (first entry)
DE Human tyrosine kinase substrate cks118/Dresh protein sequence.
XX
XX Tyrosine kinase substrate; TKs; TKs 107; GRUB; TKs 113; TKs 118; Dresh;
KW TKs 202; Src; rheumatoid arthritis; atherosclerosis; stroke; cancer;
KW autoimmune disorder; organ transplantation; myocardial infarction;
KW cardiomyopathy; renal failure; neurodegenerative disorder;
KW cardiovascular disorder; inflammatory bowel disease; multiple sclerosis;
KW asthma; osteoarthritis; psoriasis; rhinitis.
XX
XX Homo sapiens.
OS
XX WO200061750-A2.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000MO-US009277.
XX
XX 09-APR-1999; 99US-0128492P.
XX
XX (SUGEN-) SUGEN INC.
XX
XX Phan H, Courtnidge SA;
XX
XX WPI; 2000-679486/66.
XX
XX DR N-PSDB; AAC66078.
XX
XX Novel tyrosine kinase substrate polypeptides and polynucleotides, used to
PT diagnose and treat diseases such as rheumatoid arthritis,
PT atherosclerosis and cancer.
XX
XX Claim 2; Fig 7; 12pp; English.
XX
XX This invention relates to isolated, enriched or purified nucleic acid
CC molecules represented by sequences AAC66076-66079 which encode tyrosine
CC kinase substrate (TKs) TKs 107/GRUB, TKs 113, TKs 118/Dresh or TKs 202
CC represented by sequences AA85660-85663. The TKs proteins are probes
CC for cytoplasmic tyrosine kinase Src. The invention encompasses a probe
CC sequences, recombinant cells, antibodies specific for the TKs proteins,
CC and hybridomas producing the antibodies. The TKs proteins exhibit
CC antineumatic, antiarthritic, antiarteriosclerotic; immunosuppressive;
CC cardiatic; cerebroprotective; neuroprotective; cyostatic;
CC antiinflammatory; osteopathic; antipsoriatic; antiallergic; and
CC antiaesthetic activities. The tyrosine kinase substrates are used to
CC treat and diagnose diseases or disorders selected from rheumatoid
CC arthritis, atherosclerosis, autoimmune disorders, organ transplantation,
CC myocardial infarction, cardiomyopathy, stroke, renal failure, oxidative
CC stress-related neurodegenerative disorders, cardiovascular disorders,
CC cancer, and immune-related disorders selected from chronic inflammatory
CC bowel disease, chronic inflammatory pelvic disease, multiple sclerosis,
CC asthma, osteoarthritis, psoriasis, and rhinitis. The proteins can also be
CC used to identify modulators of its activity. DNA sequences encoding the
CC proteins may be used in gene therapy methods, and to produce transgenic
CC animals
XX
XX Sequence 431 AA;

Query Match 20.0%; Score 796; DB 3; Length 431;
Best Local Similarity 82.5%; Pred. No. 7.1e-73; Indels 0; Gaps 0;
Matches 151; Conservative 2; Mismatches 30;

```
QY 359 RPERPPXXRPSRTASWGTWXTXSISSPOGKLRSPFLQOLTOPETHFGREPAAIXSR 418
      |||
      249 RNEQESAVHPRIRIFQKERRAMSTSISSPOGKLRSPFLQOLTOPETHFGREPAAIXSR 308
QY 419 PRADLPAPBPAPSPPCLVQAEAEAVYEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 478
      |||
      309 PRADLPAPBPAPSPPCLVQAEAEAVYEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 368
QY 479 GLSXQGLCARALYDQADDXEISFDPENLITGIEVXXEGMMRGYGPDPGFGKXPANYVE 538
      |||
      369 GLSXQGLCARALYDQADDXEISFDPENLITGIEVDEGMWNGYGPDPGFGFPANYVE 428
QY 539 LIE 541
      |||
      429 LIE 431
Db
QY
Db
RESULT 11
AAM79569
ID AAM79569 standard; protein; 458 AA.
XX
AC AAM79569;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3215.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
XX MO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 27-APR-2000; 2000US-00560875.
XX
XX 20-JUN-2000; 2000US-00598075.
XX
XX 19-JUL-2000; 2000US-00620325.
XX
XX 01-SEP-2000; 2000US-00654936.
XX
XX 15-SEP-2000; 2000US-00663561.
XX
XX 20-OCT-2000; 2000US-00693325.
XX
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y,
XX Ma Y, Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW,
XX Xue HJ, Yang Y, Wejthman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAK52702.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 20; Page 287; 6221pp: English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
```

```
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
XX Sequence 458 AA:
QY 359 RPERPPXXRPSRTASWGTWXTXSISSPOGKLRSPFLQOLTOPETHFGREPAAIXSR 418
      |||
      276 RNEQESAVHPRIRIFQKERRAMSTSISSPOGKLRSPFLQOLTOPETHFGREPAAIXSR 335
Db
QY 419 PRADLPAPBPAPSPPCLVQAEAEAVYEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 478
      |||
      336 PRADLPAPBPAPSPPCLVQAEAEAVYEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 395
Db
QY 479 GLSXQGLCARALYDQADDXEISFDPENLITGIEVXXEGMMRGYGPDPGFGKXPANYVE 538
      |||
      396 GLSXQGLCARALYDQADDXEISFDPENLITGIEVDEGMWNGYGPDPGFGFPANYVE 455
QY 539 LIE 541
      |||
      456 LIE 458
Db
RESULT 12
ABP98850
ID ABP98850 standard; protein; 377 AA.
XX
XX ABP98850;
XX
XX 15-JUL-2003 (first entry)
XX
XX Human structural and cytoskeletal associated protein #41.
XX
XX Cytoskeletal; antiarteriosclerotic; anticonvulsant; nootropic; antianginal;
XX neuroprotective; cerebroprotective; hypotensive; cardiant; osteopathic;
XX antiinflammatory; antiarthritic; vinicide; gene therapy; human;
XX structural and cytoskeleton-associated protein; SCAP; cancer; angina;
XX atherosclerosis; epilepsy; Huntington's disease; hypertension;
XX heart failure; osteoporosis; osteoarthritis.
OS Homo sapiens.
XX
XX MO2003031940-A2.
XX
XX 17-APR-2003.
XX
XX 10-OCT-2002; 2002WO-US032851.
XX
XX 12-OCT-2001; 2001US-0328931P.
XX
XX 19-OCT-2001; 2001US-0360681P.
XX
XX 02-NOV-2001; 2001US-0343896P.
XX
XX 09-NOV-2001; 2001US-0346308P.
XX
XX 16-NOV-2001; 2001US-0332385P.
XX
XX 07-DEC-2001; 2001US-0340776P.
XX
XX 11-JAN-2002; 2002US-0347703P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Becha SD, Bhatia U, Blake JF, Borowsky ML, Burrill JD, Chang H,
XX Chavala NK, Elliott VS, Emerling BM, Forsythe TJ, Gorrard AE, Lal PG,
XX Griffin JA, Hafalila AJA, Ho A, Ison CH, Kabde AE, Khare R, Lal PG,
XX Lee S, Lee SA, Lee SY, Lehr-Wason PM, Li JX, Lindquist EA, Luo W,
XX Marquis JP, Ramkumar U, Richardson TW, Sprague WW, Swarnakar A,
XX Tang YT, Warren BA, Yang J, Yue H, Zabarjadian Y, Zheng W;
XX
XX WPI; 2003-403125/38.
XX
XX N-PSDB; ACC44337.
```

PT New human structural and cytoskeleton-associated proteins (SCAP) useful
 PT for diagnosing, treating and preventing diseases or conditions associated
 XX with the aberrant SCAP expression e.g. cancer, osteoporosis, or epilepsy.
 PS Claim 1; Page 297; 361pp; English.

XX This sequence represents a novel isolated human structural and
 CC cytoskeleton-associated protein (SCAP) polypeptide. The polypeptides and
 CC polynucleotides encoding them are useful in diagnosing, treating and
 CC preventing diseases or conditions associated with the decreased
 CC expression or over expression of SCAP, such as cell proliferative (e.g.
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
 CC disease, stroke), heart (e.g. hypertension, heart failure, angina) and
 CC skeletal muscle disorder (e.g. osteoporosis, osteoarthritis) or viral
 CC infections. These are also useful in assessing the effects of exogenous
 CC compounds on the expression of nucleic acid and amino acid sequences of
 CC SCAP. The SCAP or its fragments are useful in screening compounds for
 CC effectiveness as agonist or antagonist of the polypeptides, or in
 CC altering the expression of the target polynucleotide and compounds that
 CC specifically bind to or modulate the activity of the polypeptide. The
 CC microarray is useful in monitoring or measuring protein-protein
 CC interactions, drug-target interactions, and gene expression profiles
 XX

Sequence 377 AA:

Query Match 19.9%; Score 794; DB 6; Length 377;
 Best Local Similarity 91.9%; Pred. No. 9, 6e-73;
 Matches 148; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 381 TKSISPPQPKLRSPFLQXQLTQPEHFGREPAAXSRPADLPABEPASXPCLVQAE 440
 DB 217 TTSISSPPQPKLRSPFLQXQLTQPEHFGREPAAXSRPADLPABEPASXPCLVQAE 276
 QY 441 BEAVYEEPPXGERTFYEQPLVQOQXGSEHIDHIQGGSLXQGLCARALYDQADDXE 500
 DB 277 BEAVYEEPPXGERTFYEQPLVQOQXGSEHIDHIQGGSLXQGLCARALYDQADDXE 336
 QY 501 ISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELIE 541
 DB 337 ISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELIE 377

RESULT 13

ID AAB93895 standard; protein; 439 AA.

AC AAB93895;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:13840.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX

DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

PS Claim 8; SEQ ID NO 13840; 2537bp + Sequence listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dr primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification; where the
 CC of an oligonucleotide comprising a sequence complementary to the
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence; where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13133 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX

Sequence 439 AA:

Query Match 19.9%; Score 794; DB 4; Length 439;
 Best Local Similarity 91.9%; Pred. No. 1, 2e-72;
 Matches 148; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 381 TKSISPPQPKLRSPFLQXQLTQPEHFGREPAAXSRPADLPABEPASXPCLVQAE 440
 DB 279 TTSISSPPQPKLRSPFLQXQLTQPEHFGREPAAXSRPADLPABEPASXPCLVQAE 338

QY 441 BEAVYEEPPXGERTFYEQPLVQOQXGSEHIDHIQGGSLXQGLCARALYDQADDXE 500

DB 339 BEAVYEEPPXGERTFYEQPLVQOQXGSEHIDHIQGGSLXQGLCARALYDQADDXE 398

QY 501 ISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELIE 541

DB 399 ISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELIE 439

RESULT 14

ID ABUS2367 standard; protein; 525 AA.

AC ABUS2367;

DT 03-MAR-2003 (first entry)

DE Human GPCR related protein NOV22a.

XX Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.

OS Homo sapiens.

PN WO200279398-A2.

PD 10-OCT-2002.

PF 08-MAR-2002; 2002WO-US007355.

PR 08-MAR-2001; 2001US-0274194F.

PR 08-MAR-2001; 2001US-0274281F.

PR 08-MAR-2001; 2001US-0274322P.
 PR 09-MAR-2001; 2001US-0274849P.
 PR 13-MAR-2001; 2001US-0275578P.
 PR 13-MAR-2001; 2001US-0275579P.
 PR 13-MAR-2001; 2001US-0275601P.
 PR 14-MAR-2001; 2001US-0276000P.
 PR 16-MAR-2001; 2001US-0276776P.
 PR 19-MAR-2001; 2001US-0276994P.
 PR 20-MAR-2001; 2001US-0277239P.
 PR 20-MAR-2001; 2001US-0277332P.
 PR 21-MAR-2001; 2001US-0277791P.
 PR 22-MAR-2001; 2001US-0278339P.
 PR 23-MAR-2001; 2001US-0278152P.
 PR 26-MAR-2001; 2001US-0278894P.
 PR 27-MAR-2001; 2001US-0278999P.
 PR 30-MAR-2001; 2001US-0280239P.
 PR 02-APR-2001; 2001US-0280802P.
 PR 02-MAY-2001; 2001US-0288052P.
 PR 02-MAY-2001; 2001US-0288066P.
 PR 02-MAY-2001; 2001US-0288228P.
 PR 17-MAY-2001; 2001US-0291766P.
 PR 07-JUN-2001; 2001US-0296693P.
 PR 08-JUN-2001; 2001US-0296856P.
 PR 08-JUL-2001; 2001US-0303230P.
 PR 05-JUL-2001; 2001US-0303237P.
 PR 08-AUG-2001; 2001US-0310913P.
 PR 13-AUG-2001; 2001US-0311978P.
 PR 14-AUG-2001; 2001US-0312191P.
 PR 16-AUG-2001; 2001US-0312916P.
 PR 17-AUG-2001; 2001US-0313182P.
 PR 20-AUG-2001; 2001US-0313626P.
 PR 21-AUG-2001; 2001US-0314018P.
 PR 27-AUG-2001; 2001US-0315227P.
 PR 10-SEP-2001; 2001US-0318403P.
 PR 10-SEP-2001; 2001US-0318510P.
 PR 14-SEP-2001; 2001US-0322936P.
 PR 14-SEP-2001; 2001US-0322350P.
 PR 27-SEP-2001; 2001US-0325378P.
 PR 09-NOV-2001; 2001US-0332486P.
 PR 09-NOV-2001; 2001US-0345359P.
 PR 07-MAR-2002; 2002US-00094886.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Kekuda R, Tchernov VT, Liu X, Spytek KA, Patunajan M, Boldog FL, Bugess CE, Vernet CM, Li L, Gorman L, Malyankar UM, Taupier RJ, Miller CE, Casman SJ, Pena CE, Gangoli EA, Gusev V, Smithson G, Zerhusen BD, Gerlach V, Pochart PF, Fernandes ER, Shinkets RA, Raselli L, Spaderna SK, Larochele WJ, Zhong M, Khramtsov NV, Voss EZ, Herrmann UJ;
 PI
 XX WPI; 2003-058423/05.
 DR N-PSDB; ABX70458.
 DR
 XX
 PT NOXV polypeptides and polynucleotides, useful for treating a syndrome related to a human disease associated with the NOXV polypeptide e.g., cancer.
 PT
 XX
 XX
 PS Claim 1; Page 156; 413pp; English.
 XX
 XX The present invention relates to the isolation of novel human polypeptides referred to as NOXV (NOV1-NOV44), variants of these proteins, and the polynucleotide sequences encoding them. The NOXV proteins of the invention are G-protein coupled receptor (GPCR) related proteins. The sequences of the invention are useful in the manufacture of a medicament for treating a syndrome related to a human disease associated with the polypeptides e.g. cancer. AB052311-AB052408 represent the human NOXV proteins of the invention
 CC
 CC
 XX Sequence 525 AA;

Query Match 18.7%; Score 746.5; DB 6; Length 525;
 Best Local Similarity 67.3%; Pred. No. 1.2e-67;
 Matches 148; Conservative 9; Mismatches 34; Indels 29; Gaps 3;
 QY 381 TSSISSPQKLRSPFLQXLTPTPTFGREPAAXSRPADLPAREPAPSPCLVQAE 440
 DB 304 TTSISSPQKLRSPFLQXLTPTPTFGREPAAXSRPADLPAREPAPSPCLVQAE 363
 QY 441 EEAAYEEPEQETFEQPPVLVQOQXGSEHIDHIIQGGSLSGQGLCARLYDQADDE 500
 DB 364 EEAAYEEPEQETFEQPPVLVQOQXGSEHIDHIIQGGSLSGQGLCARLYDQADDE 423
 QY 501 ISFDPENLITGIEVXXEGMWRGYPDGHGKX-PANYVELTEX----- 542
 DB 424 ISFDPENLITGIEVIDEGWRGYPDGHGKSLPTWSSLSEAGTSCPSPLRHGFLTAG 483
 QY 543 --GXGHILKFPDQ-----TWLPYCKRRPXXHSL 571
 DB 484 RGQLGVDIQHSRRNTPSEDEASGLPAMQTPVTPNAA 523
 RESULT 15
 AA73942
 ID AA73942 standard; protein; 162 AA.
 AC
 XX AA73942;
 AC
 XX 14-MAR-2000 (first entry)
 DT
 XX Human prostate tumor EST fragment derived protein #129.
 DE
 XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
 KW treatment.
 KM Homo sapiens.
 OS
 XX DE19820190-A1.
 PN
 XX 04-NOV-1999.
 PD
 XX 28-APR-1998; 98DE-01020190.
 XX
 PF 28-APR-1998; 98DE-01020190.
 XX
 XX 28-APR-1998; 98DE-01020190.
 XX
 PR (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 XX
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 PI WPI; 1999-621386/54.
 DR N-PSDB; AA252900.
 XX
 PT New human nucleic acid sequences from pancreatic tumors, and related proteins.
 PT
 XX
 XX Claim 23; Page 363; 502pp; German.
 XX
 CC This invention describes novel polypeptides and their encoding nucleic acid derived from human pancreatic tumor tissue which have cytostatic activity. The sequences are also useful in producing pharmaceutical compositions for treatment of pancreatic tumors. AA73814-Y74252 represent protein fragments encoded by the human pancreatic tumor cDNA library derived expressed sequence tag (EST) sequences represented in
 CC AA252858-Z53014
 CC
 XX Sequence 162 AA;
 SQ
 Query Match 16.2%; Score 647; DB 2; Length 162;
 Best Local Similarity 92.2%; Pred. No. 4.5e-58;
 Matches 119; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 650 GASGRLLMPSAFLICLXFSLSKSGWPPFLFMXIGNSRRRLFLAEFTYXRVVDHGLAA 709
 DB 2 GASGRLLMPSAFLICLXFSLSKSGWPPFLFMXIGNSRRRLFLAEFTYXRVVDHGLAA 61

Qy	710	GNLSCXXLLCAPHSISLSICLGXKXWGCWPSSHPGYSKXADTTCSSTRLTRCLQXVCA	769
Db	62	GNLSCWSLLCAPHSISLSICLGXKXWGCWPSSHPGYSKXADTTCSSTRLTRCLQAPVCA	121
Qy	770	STDSDFRKS	778
Db	122	STDSDFRKS	130

Search completed: April 27, 2004, 10:52:56
Job time : 55.8053 secs

Wed Apr 28 09:57:51 2004

us-10-028-952a-9.rat

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:51:05 ; Search time 20.1141 seconds
(without alignments)
2035.354 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989
Sequence: 1 HEIPVPTVYPAKPOXXERA.....DFRKRKKKKLEKMTSSX 793

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A COMB .pep.*
2: /cgn2_6/ptodata/2/1aa/5B COMB .pep.*
3: /cgn2_6/ptodata/2/1aa/6A COMB .pep.*
4: /cgn2_6/ptodata/2/1aa/6B COMB .pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS COMB .pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1 .pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	800.5	20.1	355	4	US-08-630-915A-192
2	628.5	15.8	433	4	US-08-630-915A-18
3	409	10.3	538	4	US-08-994-076-2
4	409	10.3	538	4	US-09-643-476-2
5	300	7.5	58	4	US-08-630-915A-118
6	300	7.5	58	4	US-08-630-915A-214
7	239.5	6.0	486	2	US-08-942-423-3
8	228.5	5.7	486	2	US-08-942-423-2
9	228.5	5.7	486	2	US-08-942-423-4
10	208	5.2	546	2	US-08-630-915A-26
11	208	5.2	546	2	US-08-630-915A-14
12	185.5	4.7	57	4	US-08-630-915A-116
13	180.5	4.5	53	2	US-08-942-423-37
14	178.5	4.5	57	4	US-08-630-915A-122
15	176.5	4.4	54	2	US-08-942-423-36
16	175.5	4.4	54	2	US-08-942-423-38
17	171.5	4.3	53	2	US-08-942-423-39
18	171.5	4.3	53	2	US-08-942-423-30
19	171.5	4.3	53	2	US-08-942-423-35
20	170.5	4.3	53	2	US-08-942-423-33
21	170.5	4.3	53	2	US-08-942-423-34
22	169.5	4.2	48	3	US-08-938-830-7
23	169.5	4.2	48	3	US-08-020-222-7
24	166.5	4.2	50	3	US-08-938-830-6
25	166.5	4.2	50	3	US-09-020-222-6
26	165.5	4.1	53	2	US-08-942-423-31
27	165.5	4.1	53	2	US-08-942-423-32

28	156.5	3.9	53	2	US-08-942-423-39	Sequence 39, Appl
29	148	3.7	205	4	US-08-630-915A-16	Sequence 16, Appl
30	148	3.7	401	2	US-08-549-004A-5	Sequence 5, Appl
31	148	3.7	401	3	US-09-051-982A-5	Sequence 5, Appl
32	141	3.5	261	2	US-08-691-818B-4	Sequence 4, Appl
33	138.5	3.5	456	4	US-09-252-991A-1735	Sequence 1735, A
34	130.5	3.3	497	4	US-09-252-991A-23620	Sequence 23620, A
35	127	3.2	696	3	US-08-906-865-4	Sequence 4, Appl
36	127	3.2	696	4	US-09-129-668-4	Sequence 4, Appl
37	125.5	3.1	1706	4	US-09-252-991A-31760	Sequence 31760, A
38	123	3.1	368	4	US-08-630-915A-20	Sequence 20, Appl
39	122	3.1	123	4	US-09-107-532A-5583	Sequence 5583, Ap
40	122	3.1	123	4	US-09-107-532A-5584	Sequence 5584, Ap
41	122	3.1	123	4	US-09-134-000C-6555	Sequence 6555, Ap
42	121	3.0	181	4	US-09-252-991A-26910	Sequence 26910, A
43	120.5	3.0	59	4	US-08-630-915A-117	Sequence 117, App
44	120.5	3.0	78	4	US-09-621-976-7727	Sequence 7727, Ap
45	119.5	3.0	802	4	US-09-252-991A-25050	Sequence 25050, A

ALIGNMENTS

RESULT 1
US-08-630-915A-192
Sequence 192, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLER, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OR INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TRIEUX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-192
Query Match 20.1%, Score 800.5, DB 4, Length 355,
Best Local Similarity 64.9%, Pred. No. 5.3e-75,
Matches 157, Conservative 9, Mismatches 47, Indels 29, Gaps 1,

Db 1 MILFSTNTDRESGAS-----DLTFW-----VLSRRCQKSYH 33
QY 215 XXAEFTKRWIV-----HPLIG---NXSWDXTVRQVS-- 243
Db 34 -----RDNWLVAAKRSKRRRFLILRCRLFLSLGSRHQALDQSPPTNREREGLDRRETL 88
QY 244 FTLLMCCCHGNPAQYERNRRXXHLYVYLGGANGAKLXSVGLXLNASKESRPGXTIRQ 303
Db 89 FTLLMCCCHGNPAQYERNRRFRHLVYVLGGANGAKLPSVGLLNASKESRPGXTIRQ 148
QY 304 RRGASVGLGXPPXXRLSPAPGRPLHAPRXGRARAPPAAGTGVRCXVPFVLGNGARPER 363
Db 149 RRGASVGLGXMPVPRLSPPAGAP-----PPAPRA-----REGACPAAR 187
QY 364 PXXRPSRTARSGWTWRXTXSISSPOP 389
Db 188 RDRGPFVSALRPGKRGAGKAAPSP 213

RESULT 4
US-09-643-476-2
Sequence 2, Application US/09643476
Patent No. 6599495
GENERAL INFORMATION:
APPLICANT: EYIN, Jr., Paul R.
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE
OF MAMMASTATIN AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould P.C.
STREET: P.O. Box 2903
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402-0903
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,476
FILING DATE: 22-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027,315
FILING DATE: 03-OCT-1996
APPLICATION NUMBER: PCT/US97/18026
FILING DATE: 03-OCT-1997
APPLICATION NUMBER: 08/943,828
FILING DATE: 03-OCT-1997
APPLICATION NUMBER: 08/994,076
FILING DATE: 19-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Weaver, Karrie G.
REGISTRATION NUMBER: 43,245
REFERENCE/DOCKET NUMBER: 4273.1USC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5500
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-643-476-2

Query Match 10.3%; Score 409; DB 4; Length 538;

Best Local Similarity 41.0%; Pred. No. 5,9e-34;
Matches 109; Conservative 11; Mismatches 62; Indels 84; Gaps 6;
QY 155 LILFSTNTDXXGASRGSPFPFGYAGVKRTTGTITGLMRPSVXSDVAFSDYQSSYH 214
Db 1 MILFSTNTDRESGAS-----DLTFW-----VLSRRCQKSYH 33
QY 215 XXAEFTKRWIV-----HPLIG---NXSWDXTVRQVS-- 243
Db 34 -----RDNWLVAAKRSKRRRFLILRCRLFLSLGSRHQALDQSPPTNREREGLDRRETL 88
QY 244 FTLLMCCCHGNPAQYERNRRXXHLYVYLGGANGAKLXSVGLXLNASKESRPGXTIRQ 303
Db 89 FTLLMCCCHGNPAQYERNRRFRHLVYVLGGANGAKLPSVGLLNASKESRPGXTIRQ 148
QY 304 RRGASVGLGXPPXXRLSPAPGRPLHAPRXGRARAPPAAGTGVRCXVPFVLGNGARPER 363
Db 149 RRGASVGLGXMPVPRLSPPAGAP-----PPAPRA-----REGACPAAR 187
QY 364 PXXRPSRTARSGWTWRXTXSISSPOP 389
Db 188 RDRGPFVSALRPGKRGAGKAAPSP 213

RESULT 5
US-08-630-915A-118
Sequence 118, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 336
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9098
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-118

Query Match 7.5%; Score 300; DB 4; Length 58;
Best Local Similarity 93.1%; Pred. No. 6,9e-24;

Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 483 QGLCARALYDQADDEISFDPENLITGIEVXXEGWNRGYPDGHFGMKPANYVELI 540
DB 1 QGLCARALYDQADDEISFDPENLITGIEVDEGWMRGYGDGHFGMPANYVELI 58

RESULT 6
US-08-630-915A-214
Sequence 214, Application US/08630915A

PATENT No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mierck, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-214

Query Match 7.5%; Score 300; DB 4; Length 58;
Best Local Similarity 93.1%; Pred. No. 6.9e-24;
Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 483 QGLCARALYDQADDEISFDPENLITGIEVXXEGWNRGYPDGHFGMKPANYVELI 540
DB 1 QGLCARALYDQADDEISFDPENLITGIEVDEGWMRGYGDGHFGMPANYVELI 58

RESULT 7
US-08-942-423-3
Sequence 3, Application US/08942423

PATENT No. 5891673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro
APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Lck Binding Protein
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:

ADDRESSEE: Syntex (U.S.A.) Inc.
STREET: 3401 Hillyview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: LCK BINDING PROTEIN
US-08-942-423-3

Query Match 6.0%; Score 239.5; DB 2; Length 486;
Best Local Similarity 28.1%; Pred. No. 2.4e-16;
Matches 81; Conservative 20; Mismatches 88; Indels 99; Gaps 11;

QY 290 ASKESRPKXITRRRQASVGLGKPKXRLPPRAGRPLHAPRRGRARAPRAGTGRGX 349
DB 262 ARQOERKAVTKRREAPQ-----PVIAEPPAVPAVL--PKIISSEAMP----- 305
QY 350 VPFVINGARPERPRPXXRPSRTASWGTWXTXSISSPOKURSPTLOXQLOPETHFG 409
DB 306 -----VQTPSSR--SEPVATISREHPV-----FLPPIRQLPEDN-- 338
QY 410 REPAAKXSRPADIPAEPPAPKPCCL--VQAEAEVYEEPRQEFTFYQPLVQOQXX 466
DB 339 -----BEP-PALPPTLLEGLQVEEPEVYEAEPPEP--BPEPEPDYE 379
QY 467 GSEHIDHHIQ-----GQGUSXQGLCARALYDY 493
DB 380 DVEENDRHEQDEDEGDYEVLPEPDESSFSALAGSSGCPAGAGAGVALGISVALYDY 439
QY 494 QAADDEISFDPENLITGIEVXXEGWNRGYPDGHFGMKPANYVELI 541
DB 440 QEGSDDEISFDPDVIDITDIEMVDEGWMRG-RCHGHFGLFPANYVELI 486

RESULT 8
US-08-942-423-2
Sequence 2, Application US/08942423

PATENT No. 5891673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro
APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Lck Binding Protein
NUMBER OF SEQUENCES: 68

```

CORRESPONDENCE ADDRESS:
ADDRESS: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: HSI
US-08-942-423-2

Query Match      5.7%; Score 228.5; DB 2; Length 486;
Best Local Similarity 29.5%; Pred. No. 3.3e-15;
Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GANG-AKLXSVGLKLNASKSESPPGXTIRQ--RGASVGLGXPKXRLSPAGRP--PLH 328
DB 234 GARGLKAFKESLAERKREBEERKQOMARQOERKAVVMKSRREVQOPSMPEVEBPAPAQ 293
QY 329 APRXRGARAPPRAGTGVACVFPVLGNGARPERRPXKPSRTARSWGTWXTXISSPQ 388
DB 294 LPKTISSEVWP-----AESHLPPESQPVRSRREY-----PV 325
QY 389 PG-KLRSPFLQXQLTQPEHFGREPAAXSR-PRADLPAEBPABSPCLVQAEBAAYE 446
DB 326 PSLPTQSPQLQNHLEDN-----EPALPFRTPBGLQVEBEPVYEAAPBL-EPREBPYE 379
QY 447 EPXBEFTYEPQPPVVOQXGXSEHIDHNI-----QGQGLSXQGLCARALYD 492
DB 380 PEPEFTPEYEDVGEIDRQDEDAEGDYEDVLEPEDTPSLSYQAGPSAGAGAGISALALYD 439
QY 493 YQAADXKISFPDENLITGIEVXXEGWWRGYGPDGHFGKXPANYVELI 540
DB 440 YQEGSDLSFDPDDIITIDEMVDEGWWRGQ-CRGHFGLFPANYVKLL 486

```

```

RESULT 9
US-08-630-915A-26
Sequence 26, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, NO. 6309820H
APPLICANT: KAY, Brian K.

```

```

APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TEXT: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-26

Query Match      5.7%; Score 228.5; DB 4; Length 486;
Best Local Similarity 29.5%; Pred. No. 3.3e-15;
Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GANG-AKLXSVGLKLNASKSESPPGXTIRQ--RGASVGLGXPKXRLSPAGRP--PLH 328
DB 234 GARGLKAFKESLAERKREBEERKQOMARQOERKAVVMKSRREVQOPSMPEVEBPAPAQ 293
QY 329 APRXRGARAPPRAGTGVACVFPVLGNGARPERRPXKPSRTARSWGTWXTXISSPQ 388
DB 294 LPKTISSEVWP-----AESHLPPESQPVRSRREY-----PV 325
QY 389 PG-KLRSPFLQXQLTQPEHFGREPAAXSR-PRADLPAEBPABSPCLVQAEBAAYE 446
DB 326 PSLPTQSPQLQNHLEDN-----EPALPFRTPBGLQVEBEPVYEAAPBL-EPREBPYE 379
QY 447 EPXBEFTYEPQPPVVOQXGXSEHIDHNI-----QGQGLSXQGLCARALYD 492
DB 380 PEPEFTPEYEDVGEIDRQDEDAEGDYEDVLEPEDTPSLSYQAGPSAGAGAGISALALYD 439
QY 493 YQAADXKISFPDENLITGIEVXXEGWWRGYGPDGHFGKXPANYVELI 540
DB 440 YQEGSDLSFDPDDIITIDEMVDEGWWRGQ-CRGHFGLFPANYVKLL 486

```

```

RESULT 10
US-08-942-423-4
Sequence 4, Application US/08942423
Patent No. 5891673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro
APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Lock Binding Protein
NUMBER OF SEQUENCES: 68

```

CORRESPONDENCE ADDRESS:
ADDRESSSEE: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 496-3529
TELEFAX: (415) 852-1698
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE: HSI
ORGANISM: HSI
US-08-942-423-4

Query Match 5.2%; Score 208; DB 2; Length 546;
Best Local Similarity 35.6%; Pred. No. 5.3e-13;
Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;

QY 386 SPQPGKLRSPFLQXLTQPTTHGCRPAXXSRPADLPABEPAPSPCTLVQAESEAVY 445
DB 407 SPQIEDRPP--SSPIEDAPFKAEPSYRGSEP-----EPEYS-----IEAGIP 450

QY 446 EEPXEOETFEYQPPVQOQXGSEHIDHIIQGG-----LSXQGLCARALYDQADDX 499
DB 451 EASQOGLTYTSEVYE-----TTEAPGHYQAEDDTYDGYESDLGITATLYDQAGDD 505

QY 500 EISFDPENLITGIEVXXEGWGRGYPDGHGKMPANVEL 539
DB 506 EISFDPDDITITNEMIDDGWGRGV-CKGRYGLFPANVEL 544

RESULT 11
US-08-630-915A-14
Sequence 14, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, NO. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-14

Query Match 5.2%; Score 208; DB 4; Length 546;
Best Local Similarity 35.6%; Pred. No. 5.3e-13;
Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;

QY 386 SPQPGKLRSPFLQXLTQPTTHGCRPAXXSRPADLPABEPAPSPCTLVQAESEAVY 445
DB 407 SPQIEDRPP--SSPIEDAPFKAEPSYRGSEP-----EPEYS-----IEAGIP 450

QY 446 EEPXEOETFEYQPPVQOQXGSEHIDHIIQGG-----LSXQGLCARALYDQADDX 499
DB 451 EASQOGLTYTSEVYE-----TTEAPGHYQAEDDTYDGYESDLGITATLYDQAGDD 505

QY 500 EISFDPENLITGIEVXXEGWGRGYPDGHGKMPANVEL 539
DB 506 EISFDPDDITITNEMIDDGWGRGV-CKGRYGLFPANVEL 544

RESULT 12
US-08-630-915A-116
Sequence 116, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, NO. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 116:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-116

Query Match 4.7%; Score 185.5; DB 4; Length 57;
Best Local Similarity 62.5%; Pred. No. 5.8e-12;
Matches 35; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

QY 484 GLCARALDYQAADDXISFPDENLITIGIEVXXEGMWRGYGPDGHFGMXKPNANYVEL 539
DB 2 GISAIALVDYQAGSDDELSPDDITITIDENVDEGMWRGQ-CRGHFGLPFANYVEL 56

RESULT 13
US-08-942-423-37
Sequence 37, Application US/08942423
Patent No. 5891673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro
APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Lock Binding Protein
NUMBER OF SEQUENCES: 68
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 853-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-942-423-37

Query Match 4.5%; Score 180.5; DB 2; Length 53;
Best Local Similarity 63.0%; Pred. No. 1.7e-11;
Matches 34; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

QY 487 ARAALDYQAADDXISFPDENLITIGIEVXXEGMWRGYGPDGHFGMXKPNANYVELI 540
DB 1 ARAALDYQAAGSDDELSPDDITITIDENVDEGMWRGQ-CRGHFGLPFANYVELL 53

RESULT 14
US-08-630-915A-122
Sequence 122, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOMKES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-122

Query Match 4.5%; Score 178.5; DB 4; Length 57;
Best Local Similarity 57.9%; Pred. No. 3.1e-11;
Matches 33; Conservative 9; Mismatches 14; Indels 1; Gaps 1;

QY 484 GLCARALDYQAADDXISFPDENLITIGIEVXXEGMWRGYGPDGHFGMXKPNANYVELI 540
DB 2 GISAIALVDYQAGSDDELSPDDITITIDENVDEGMWRGQ-CRGHFGLPFANYVELL 57

RESULT 15
US-08-942-423-36
Sequence 36, Application US/08942423
Patent No. 5891673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro

Wed Apr 28 09:57:51 2004

us-10-028-952a-9.rat

Page 8

APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Tick Binding Protein
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 54 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-942-423-36

Query Match 4.4%; Score 176.5; DB 2; Length 54;
Best Local Similarity 60.0%; Pred. No. 4.7e-11;
Matches 33; Conservative 8; Mismatches 13; Indels 1; Gaps 1;
QY 487 ARALYDYOADDXEISFDPENLITGIEVXXEGWRCGSPDGHFGKXPPATYVELIE 541
DB 1 AIALYDYOQSGSDLEISFPDDITTDIEWVEGWRGQ-CRHHGLFPATYKLE 54

Search completed: April 27, 2004, 10:56:53
Job time: 21.1141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: April 27, 2004, 10:55:20 / Search time 40.7311 Seconds

(without alignments)
5382.733 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989

Sequence: 1 HEIPVPTVYPAKPOXHERA.....DFRKSXXXXKLEKLTWISX 793

Scoring table:

BLOSUM62

Searched: 113595 seqs, 276475211 residues

Total number of hits satisfying chosen parameters: 113595

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3824	95.9	793	14	US-10-028-952A-9
2	1288	33.3	784	14	US-10-028-952A-10
3	800.5	20.1	355	9	US-09-879-957-132
4	746.5	18.7	525	15	US-10-094-886-114
5	628.5	15.8	433	9	US-09-879-957-18
6	300	7.5	38	9	US-09-879-957-118
7	300	7.5	58	9	US-09-879-957-214
8	241	6.0	55	9	US-09-925-299-1334
9	241	6.0	55	10	US-09-925-299-1334
10	226.5	5.7	486	9	US-09-879-957-26
11	226.5	5.7	643	15	US-10-369-493-5931
12	208	5.2	546	9	US-09-879-957-14
13	192	4.8	40	9	US-09-925-299-1277
14	192	4.8	40	10	US-09-925-299-1277
15	190	4.8	769	15	US-10-369-493-4212

16	185.5	4.7	57	9	US-09-879-957-116	Sequence 116, App
17	178.5	4.5	57	9	US-09-879-957-122	Sequence 122, App
18	165	4.1	57	14	US-10-083-357-1137	Sequence 1137, App
19	161	4.0	57	14	US-10-083-357-1147	Sequence 1147, App
20	148	3.7	205	9	US-09-879-957-16	Sequence 16, App
21	145	3.6	54	9	US-09-925-299-1314	Sequence 1314, App
22	145	3.6	54	10	US-09-925-299-1314	Sequence 1314, App
23	145	3.6	54	10	US-09-764-891-4667	Sequence 4667, App
24	143	3.6	592	15	US-10-369-493-1534	Sequence 1534, App
25	142.5	3.6	97	9	US-09-925-299-1071	Sequence 1071, App
26	142.5	3.6	97	10	US-09-925-299-1071	Sequence 1071, App
27	141	3.5	261	14	US-10-247-671-134	Sequence 134, App
28	141	3.5	261	15	US-10-435-666-27	Sequence 27, App
29	139.5	3.5	19608	15	US-10-084-846A-9	Sequence 8, App
30	138.5	3.5	52	14	US-10-083-357-1128	Sequence 1148, App
31	138.5	3.5	19652	15	US-10-084-846A-7	Sequence 7, App
32	138	3.5	50	9	US-09-925-299-1313	Sequence 1313, App
33	138	3.5	50	10	US-09-981-151A-92	Sequence 1133, App
34	138	3.5	59	11	US-10-084-846A-5	Sequence 5, App
35	135.5	3.4	19723	15	US-10-083-357-1133	Sequence 103, App
36	135	3.4	30	14	US-10-042-865-103	Sequence 6, App
37	135	3.4	57	12	US-10-084-846A-6	Sequence 4, App
38	130	3.3	19662	15	US-10-202-481-4	Sequence 26, App
39	129.5	3.2	839	15	US-10-336-472-26	Sequence 66, App
40	129.5	3.2	1096	12	US-10-236-417-66	Sequence 10, App
41	129.5	3.2	1098	12	US-10-362-571-10	Sequence 2, App
42	129.5	3.2	1098	15	US-10-202-481-2	Sequence 12877, App
43	129.5	3.2	1098	15	US-10-369-493-12877	Sequence 8, App
44	128.5	3.2	174	15	US-09-767-215-8	
45	127.5	3.2	58	9	US-09-767-215-8	

ALIGNMENTS

RESULT 1

US-10-028-952A-9

Sequence 9, Application US/10028952A

Publication No. US20030157576A1

GENERAL INFORMATION:

APPLICANT: Ertvin, Jr., Paul R.

TITLE OR INVENTION: EPITHELIAL CELL GROWTH INHIBITORS

FILE REFERENCE: 4273.3USMT

CURRENT FILING DATE: 2001-12-18

PRIOR APPLICATION NUMBER: PCT/US00/16900

PRIOR FILING DATE: 2000-06-19

PRIOR APPLICATION NUMBER: 60/139,995

PRIOR FILING DATE: 1999-06-18

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 9

LENGTH: 793

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: UNSURE

LOCATION: (16)..(17)

OTHER INFORMATION: Xaa is any amino acid or may be absent

FEATURE:

NAME/KEY: UNSURE

LOCATION: (42)

OTHER INFORMATION: Xaa is any amino acid or may be absent

FEATURE:

NAME/KEY: UNSURE

LOCATION: (45)

OTHER INFORMATION: Xaa is any amino acid or may be absent

FEATURE:

NAME/KEY: UNSURE

LOCATION: (49)

OTHER INFORMATION: Xaa is any amino acid or may be absent

FEATURE:

NAME/KEY: UNSURE

```

1 OTHER INFORMATION: Xaa is any amino acid or may be absent
2 FEATURE:
3 NAME/KEY: UNSURE
4 LOCATION: (274)
5 OTHER INFORMATION: Xaa is any amino acid or may be absent
6 FEATURE:
7 NAME/KEY: UNSURE
8 LOCATION: (282)
9 OTHER INFORMATION: Xaa is any amino acid or may be absent
10 FEATURE:
11 NAME/KEY: UNSURE
12 LOCATION: (287)
13 OTHER INFORMATION: Xaa is any amino acid or may be absent
14 FEATURE:
15 NAME/KEY: UNSURE
16 LOCATION: (299)
17 OTHER INFORMATION: Xaa is any amino acid or may be absent
18 FEATURE:
19 NAME/KEY: UNSURE
20 LOCATION: (312)
21 OTHER INFORMATION: Xaa is any amino acid or may be absent
22 FEATURE:
23 NAME/KEY: UNSURE
24 LOCATION: (315) - (316)
25 OTHER INFORMATION: Xaa is any amino acid or may be absent
26 FEATURE:
27 NAME/KEY: UNSURE
28 LOCATION: (332)
29 OTHER INFORMATION: Xaa is any amino acid or may be absent
30 FEATURE:
31 NAME/KEY: UNSURE
32 LOCATION: (349)
33 OTHER INFORMATION: Xaa is any amino acid or may be absent
34 FEATURE:
35 NAME/KEY: UNSURE
36 LOCATION: (365) - (366)
37 OTHER INFORMATION: Xaa is any amino acid or may be absent
38 FEATURE:
39 NAME/KEY: UNSURE
40 LOCATION: (380)
41 OTHER INFORMATION: Xaa is any amino acid or may be absent
42 FEATURE:
43 NAME/KEY: UNSURE
44 LOCATION: (382)
45 OTHER INFORMATION: Xaa is any amino acid or may be absent
46 FEATURE:
47 NAME/KEY: UNSURE
48 LOCATION: (399)
49 OTHER INFORMATION: Xaa is any amino acid or may be absent
50 FEATURE:
51 NAME/KEY: UNSURE
52 LOCATION: (415) - (416)
53 OTHER INFORMATION: Xaa is any amino acid or may be absent
54 FEATURE:
55 NAME/KEY: UNSURE
56 LOCATION: (432)
57 OTHER INFORMATION: Xaa is any amino acid or may be absent
58 FEATURE:
59 NAME/KEY: UNSURE
60 LOCATION: (449)
61 OTHER INFORMATION: Xaa is any amino acid or may be absent
62 FEATURE:
63 NAME/KEY: UNSURE
64 LOCATION: (465) - (466)
65 OTHER INFORMATION: Xaa is any amino acid or may be absent
66 FEATURE:
67 NAME/KEY: UNSURE
68 LOCATION: (482)
69 OTHER INFORMATION: Xaa is any amino acid or may be absent
70 FEATURE:
71 NAME/KEY: UNSURE
72 LOCATION: (499)
73 OTHER INFORMATION: Xaa is any amino acid or may be absent

```

```

FEATURE:
NAME/KEY: UNSURE
LOCATION: (515)..(516)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (532)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (542)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (544)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

Query Match 95.9%; Score 3824; DB 14; Length 793;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 792; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 HEIPVPTYYPAKPOXERAMENQKKTLLSTLVHGBETXVXNKKXAPGAPVPSRG 60
DB 1 HEIPVPTYYPAKPOXERAMENQKKTLLSTLVHGBETXVXNKKXAPGAPVPSRG 60
QY 61 AAGGXXRPGCPVVKTHYSRFTDVRRGGEPRGALAXARPARRGATRSCKXAWG 120
DB 61 AAGGXXRPGCPVVKTHYSRFTDVRRGGEPRGALAXARPARRGATRSCKXAWG 120
QY 121 VILGAYTCQTVQVXGELREDRNLPMXRAKARILLFSTITDXXSASSFXPFQFXA 180
DB 121 VILGAYTCQTVQVXGELREDRNLPMXRAKARILLFSTITDXXSASSFXPFQFXA 180
QY 181 GXVRKVTGTIGLWMPVSXSDVAFYSFDVGSYSHKXAEFTKMTVHPLIGXSMXTYVR 240
DB 181 GXVRKVTGTIGLWMPVSXSDVAFYSFDVGSYSHKXAEFTKMTVHPLIGXSMXTYVR 240
QY 241 QVSTFLMKCCCHGNPAQYERNRXXHLVYVVGXGANGAKLSVGLXNASKSESPGXT 300
DB 241 QVSTFLMKCCCHGNPAQYERNRXXHLVYVVGXGANGAKLSVGLXNASKSESPGXT 300
QY 301 IQRRGASVGLGXPXXRLSPAGRPPLHAPRYRGARAPRAAGTGVRCXVPVLGNARP 360
DB 301 IQRRGASVGLGXPXXRLSPAGRPPLHAPRYRGARAPRAAGTGVRCXVPVLGNARP 360
QY 361 ERPRXXRPTARMSGTWXTXISPOGKLRSPFLQXOLQPTHTHGRBPAAXSRER 420
DB 361 ERPRXXRPTARMSGTWXTXISPOGKLRSPFLQXOLQPTHTHGRBPAAXSRER 420
QY 421 ADLPABEPASXPCLVQAEAEVYEEPEXQETFEYQEPPLVQQXXGSEHIDHIIQGG 480
DB 421 ADLPABEPASXPCLVQAEAEVYEEPEXQETFEYQEPPLVQQXXGSEHIDHIIQGG 480
QY 481 SYQGLCARALYIOAADXELISFPDENLITGIEVXXEGEMRGYGDGHFQXKPNYVLI 540
DB 481 SYQGLCARALYIOAADXELISFPDENLITGIEVXXEGEMRGYGDGHFQXKPNYVLI 540
QY 541 EXGKGHILKFPQOTMLPYCMKRRPXXHSHALFOEADPOXGKTLRAPSGLASACHPKCN 600
DB 541 EXGKGHILKFPQOTMLPYCMKRRPXXHSHALFOEADPOXGKTLRAPSGLASACHPKCN 600
QY 601 GLVPTHPSCIPRPXXTAWLPLTGYXAPCKWBPBEMPLPSGCGSGSXGASGRMLPSA 660
DB 601 GLVPTHPSCIPRPXXTAWLPLTGYXAPCKWBPBEMPLPSGCGSGSXGASGRMLPSA 660
QY 661 FICLXXFSLASGWWPPLFRXKLGNSERRELFLAFVTVXVRDHDGLAAGNLSCKXLLCA 720
DB 661 FICLXXFSLASGWWPPLFRXKLGNSERRELFLAFVTVXVRDHDGLAAGNLSCKXLLCA 720
QY 721 PHSISLCLGXGKMGCRMPSSHGYSKXADITTCSTRLTRCLQXXVCASTDSPRKSXK 780
DB 721 PHSISLCLGXGKMGCRMPSSHGYSKXADITTCSTRLTRCLQXXVCASTDSPRKSXK 780

```

```

DB 721 PHSISLCLGXGKMGCRMPSSHGYSKXADITTCSTRLTRCLQXXVCASTDSPRKSXK 780
QY 781 KXKKEKWTSS 792
DB 781 KXKKEKWTSS 792

```

RESULT 2

```

US-10-028-952a-10
Sequence 10, Application US/10028952A
Publication No. US20030157576A1
GENERAL INFORMATION:
APPLICANT: Ervin, Jr., Paul R.
TITLE OF INVENTION: EPITHELIAL CELL GROWTH INHIBITORS
FILE REFERENCE: 4273.3USM1
CURRENT APPLICATION NUMBER: US/10/028, 952A
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: PCT/US00/16900
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 60/139, 995
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 784
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (10)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (13)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (16)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (33)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (49)..(50)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (66)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (83)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (90)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (99)..(100)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (104)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (116)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

LOCATION: (133)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (142)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (147)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (149)..(150)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (166)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (173)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (183)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (199)..(200)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (204)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (216)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (233)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (242)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (249)..(250)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (255)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (266)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (281)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (283)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (299)..(300)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (316)

OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (333)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (349)..(350)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (366)..(368)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (378)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (382)..(383)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (386)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (398)..(401)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (416)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (427)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (433)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (449)..(450)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (456)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (466)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (483)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (499)..(500)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

Query Match 32.3%; Score 1288; DB 14; Length 784;
Best Local Similarity 80.3%; Pred. No. 2.7e-104;
Matches 282; Conservative 7; Mismatches 58; Indels 4; Gaps 2;

QY 37 HGEETXEYXNKMXPAPVSPRGARGXKXPCGPPYKTHYSDFXIDPVRGGEPRGAL 96
DB 5 HGEETXEYXNKMXPAPVSPRGARGXKXPCGPPYKTHYSDFXIDPVRGGEPRGAL 64
QY 97 ASXAKRPAPRPAGATRSXKXKMGVILGRYTCQTVXQVSGXELERDLFWSKAKRLI 156

Db 65 AKGAKRPAARPGATRGSGXASARWGVXLGRYTCQTXOVASXGEIENEDNLPWXRRAKALI 124
Qy 157 LIFSTNDXXSGARSFPFGFXAGVYKVTGT:GLMRPVSXDVAFXSPDVSSVHXH 216
Db 125 LIFSTNDXESGARSFPFGFXAXXVAKVTGT:GLMRPVSXDVAFXSPDVSSVHXH 184
Qy 217 AEFTRMIVHPLIGXSWDXTVVQVSFTLLMCCCHGNPAQYERNRXXHLVYLIGXA 276
Db 185 AEFTRMIVHPLIGXSWDXTVVQVSFTLLMCCCHGNPAQYERNRXXHLVYLIGXA 244
Qy 277 NGAKLXSVGLXNLNKSSESXPGTIRORGNASVGLGXPKYKRLSPAPRPLHAPRXGSA 336
Db 245 NGAKKXSVGLXNLNKSSESXPGTIRORGNASVGLGXPKYKRLSPAPRPPSTXXKAG 304
Qy 337 RAPPRA--GTGVRXCVFVLNGARPPRRPXRSRTASWGTWXTXSI 384
Db 305 RVPRRAPPGSXACPSMWTGRG-RKCGXPLAHNAHVBARAFXXSSTI 354

RESULT 3

US-09-879-957-192
Sequence 192, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
HOFMAN, No. US20020034755A1H

KAY, Brian K.
FOWLES, Dana M.

MC CONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Penile & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNMB
INFORMATION FOR SEQ ID NO: 192:

SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid

STRANDEDNESS: <Unknown>
TOPOLOGY: unknown

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 192:

US-09-879-957-192

Query Match 20.1%; Score 800.5; DB 9; Length 355;
Best Local Similarity 64.9%; Pred. No. 7,3e-62;

Matches 157; Conservative 9; Mismatches 47; Indels 29; Gaps 1;
Qy 359 RPERPXXRPSRTASWGTWXTXSISSPQKLSNPLQXLTQPTHTFGREPAAXSR 418
Db 112 RNEQSAVHREIPIQKERAMSTSISSPQKLSNPLQXLTQPTHTFGREPAAXSR 171
Qy 419 PRADLPAPERAPSTPPCLVQAEBAVYEPPEQETFEQPPVQQGAGSEHIDHIQSQ 478
Db 172 PRADLPAPERAPSTPPCLVQAEBAVYEPPEQETFEQPPVQQGAGSEHIDHIQSQ 231
Qy 479 GLSXGCLCARALYDQADDXEISFPDENLITGIEVXXEGWWRGYGPDGHPGXDPANYE 538
Db 232 GLSXGCLCARALYDQADDXEISFPDENLITGIEVXXEGWWRGYGPDGHPGXDPANYE 291
Qy 539 LIEXGKHTLXFP-----SQTWLYCWKRRPXKXHS 569
Db 292 LIDEAGTSCPSPLRNGFLIAGRGGLVDIQHSSNRRTPSDEAGLPPAMQTQVTPNA 351
Qy 570 AL 571
Db 352 AM 353

RESULT 4

US-10-094-886-114
Sequence 114, Application US/10094886
Publication No. US20040002120A1
GENERAL INFORMATION:

APPLICANT: Kekuda, Ramesh
Tchernev, Velizar T.

APPLICANT: Liu, Xiaohong
Spytek, Kimberly A.

APPLICANT: Patnursajan, Meera
Burgess, Catherine

APPLICANT: Verne, Corine A.
Li, Li

APPLICANT: Goeman, Linda
Malyankar, Uriel M.

APPLICANT: Boldog, Ferenc
Guo, Xiaojia

APPLICANT: Shenoy, Suresh
Radigaru, Muralidhara

APPLICANT: Taupier, Raymond J., Jr.
Miller, Charles

APPLICANT: Casman, Stacie
Pena, Carol

APPLICANT: Gangoli, Esha
Gusev, Vladimir

APPLICANT: Smithson, Glenda
Zerhusen, Bryan

APPLICANT: Gerlach, Valerie
Pochart, Pascal

APPLICANT: Fernandes, Elma
Shimkets, Richard

APPLICANT: Rastelli, Luca
Spaderna, Steven

APPLICANT: Laroche, William
Zhong, Mei

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-290 B
CURRENT APPLICATION NUMBER: US/10/094,886

PRIOR APPLICATION NUMBER: 60/274,322
PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 60/313,182
PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: 60/288,052
PRIOR FILING DATE: 2001-05-02

PRIOR APPLICATION NUMBER: 60/318,510
PRIOR FILING DATE: 2001-09-10

PRIOR APPLICATION NUMBER: 60/274,281
PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 60/314,018

Wed Apr 28 09:57:52 2004

us-10-028-952a-9.rapb

Page 6

PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: 60/274,194
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/274,849
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/296,693
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/313,626
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 298
SOFTWARE: Patent 2.1
SEQ ID NO: 114
LENGTH: 525
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-886-114

Query Match 18.7%; Score 746.5; DB 15; Length 525;
Best Local Similarity 67.3%; Pred. No. 6,3e-57;
Matches 148; Conservative 9; Mismatches 34; Indels 29; Gaps 3;

QY 361 TXXISSPOGKLRSPFLQXLTQPEHFGREPAAXSRPADLPAEAPASXPCLVQAE 440
DB 304 TTSISSPOGKLRSPFLQXLTQPEHFGREPAAXSRPADLPAEAPASXPCLVQAE 363
QY 441 EAAVEEPEEETFEOPPLVQOQXGSEHIDHIIQOGLSGOGLCARALYDQAADYE 500
DB 364 EAAVEEPEEETFEOPPLVQOQXGSEHIDHIIQOGLSGOGLCARALYDQAADYE 423
QY 501 ISFDENITGIEVYXEGMWRGYPGDHFGMX-PANVELIEX----- 542
DB 424 ISFDENITGIEVYXEGMWRGYPGDHFGMX-PANVELIEX----- 542
QY 543 --GXGHLXFPDQ-----TWLPCYCKRRPXXHSL 571
DB 484 RGGIGVDIQHSSRNRTSEDEASGLPPAWOTQPTVNAAM 523

RESULT 5

US-09-879-957-18
Sequence 18, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
McCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 669-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-879-957-18

Query Match 15.8%; Score 628.5; DB 9; Length 433;
Best Local Similarity 58.7%; Pred. No. 1.2e-46;
Matches 131; Conservative 12; Mismatches 43; Indels 37; Gaps 5;

QY 336 AAPPAGTGVRCVFLGNGARPERRPXXSPRTASGCT-----W 378
DB 231 AGAPSRG-----EPQEA-----SRTQWESAGCAAPRRIFFQXERA 272
QY 379 RXTXISSPOGKLRSPFLQXLTQPEHFGREPAAXSRPADLPAEAPASXPCLVQ 438
DB 273 MSTSVTSPOGKLRSPFLQXLTQPEHFGREPAAXSRPADLPAEAPASXPCLVQ 330
QY 439 AAEAAVEEPEEETFEOPPLVQOQXGSEHIDHIIQOGLSGOGLCARALYDQAAD 498
DB 331 TEEPTVEEPEEETFEOPPLVQOQXGSEHIDHIIQOGLSGOGLCARALYDQAAD 390
QY 499 XEISFDENITGIEVYXEGMWRGYPGDHFGMX-PANVELIE 541
DB 391 TEISFDENITGIEVYXEGMWRGYPGDHFGMX-PANVELIE 433

RESULT 6

US-09-879-957-118
Sequence 118, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
McCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mirock, S. Leslie

REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 118:
US-09-879-957-118

Query Match 7.5%; Score 300; DB 9; Length 58;
Best Local Similarity 93.1%; Pred. No. 8,6e-19;
Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 483 QGLCARALDYQADADDEISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELI 540
Db 1 QGLCARALDYQADADDEISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELI 58

RESULT 7
US-09-879-957-214
Sequence 214, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLER, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSES: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 214:
US-09-879-957-214

Query Match 7.5%; Score 300; DB 9; Length 58;
Best Local Similarity 93.1%; Pred. No. 8,6e-19;
Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 483 QGLCARALDYQADADDEISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELI 540
Db 1 QGLCARALDYQADADDEISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELI 58

RESULT 8
US-09-925-299-1334
Sequence 1334, Application US/09925299
Patent No. US2002005627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1334
LENGTH: 55
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (10)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (49)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (52)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (54)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (55)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1334

Query Match 6.0%; Score 241; DB 9; Length 55;
Best Local Similarity 90.4%; Pred. No. 1,2e-13;
Matches 47; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 69 CGPVTXHSYSDRFTDPRVGRGPRGALASXAPRPARRGATRSQXAXWG 120
Db 2 CGPVTXHSYSDRFTDPRVGRGPRGALASGAPRPARRGATRSQXAXWG 53

RESULT 9
US-09-925-299-1334
Sequence 1334, Application US/09925299
Patent No. US2002005627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1556
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO: 1334
 LENGTH: 55
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (10)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (49)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (52)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (54)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (55)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-299-1334

Query Match
 Best Local Similarity 90.4%; Pred. No. 1.2e-13;
 Matches 47; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 69 CDPPTKHYSDRFXTDPRRGEPGRGALASAKRPPARPPATSGXXARNG 120
 DB 2 CDPPTKHYSDRFXTDPRRGEPGRGALASAKRPPARPPATSGXXARNG 53

RESULT 10
 US-09-879-957-26
 Sequence 26, Application US/09879957
 Patent No. US20020034755A1
 GENERAL INFORMATION:
 APPLICANT: SPARKS, Andrew B.
 HOFFMAN, No. US20020034755A1h
 KAY, Brian K.
 FOMKES, Dana M.
 MCCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 USING SAME
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2721
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/879,957
 FILING DATE: 13-Jun-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/630,915
 FILING DATE: 03-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Mastrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 486 amino acids
 TYPE: amino acid
 STRANDEDNESS: <Unknown>
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 26:
 US-09-879-957-26

Query Match
 Best Local Similarity 25.5%; Pred. No. 1.1e-11;
 Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GARG-ALKXSVGLXNLASKSESRPGXTIROR--RGASVGLGXFXRLSPPARP--PLH 328
 DB 224 GARGLKXKFEELAEKREBEBAQOMARQQQERAVYMSREVOQPMVEEPAPAPQ 293
 QY 329 APRXRGARAPPRAGTGVRCXVFLVGNARPRRRXXRPSRTASWGTRXTXISRPQ 368
 DB 294 LPKTISEWVP-----AESHLPESSQPVSRREY-----PV 325
 QY 369 PG-KLRSPFLQXQLOPETHFGREPAAXSR--PRADLPAREPAPSPCLVQAEBAVYE 446
 DB 326 PSLPTRQSPQNHLENE-----EPALPRTPEGLQVBEFYEAFEL--EPEPEPVE 379
 QY 447 EPXQETFPYQPLVQCOXXGSEHIDHT-----QGQGLXQGLCARAYD 492
 DB 380 PEPETPEYVEGELDQEDDAEGYEDVLEPEDTSLSYQAGPSAGAGAGISALATYD 439
 QY 493 YQADDEXISFDPENLITGIEVXXEGWWRGYPGDGHFGMXPANYVELI 540
 DB 440 YQGESDELSPDPDITITDIEMVDEGWRGQ--CRGHGFLPANYVELL 486

RESULT 11
 US-10-369-493-5931
 Sequence 5931, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xiandeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 5931
 LENGTH: 643
 TYPE: PRT
 ORGANISM: Caenorhabditis elegans
 US-10-369-493-5931

Query Match
 Best Local Similarity 27.5%; Pred. No. 3.8e-11;
 Matches 82; Conservative 20; Mismatches 75; Indels 121; Gaps 13;

QY 319 SPFAGRPDLAPRXRGARAPPRAGTGVRCXVFLVGNARPRRRXXR-----PS--- 369
 DB 391 SKPSG-FVLKPKQVNGSPKXMPVGT-----TSPREPVNRLTEPDESTYT 436
 QY 370 -----RTASWGTRX-----TXSISQPGKLSPP-PL 397
 DB 437 PKPLAYEPPEMYKPEEMKNAVSYDAYEPPAPAPPTLATPTVIAEPPEPTAPASRY 496
 QY 398 QXQLOPETHFGREPA--XXSR-----RADLPA----- 425

Db 497 ASQYAPFVHSEFEEVAVPVSAPSHYASQYDAPPEPIDSHSSSSQLPHAHASQYDMPVY 556
QY 426 -EEP--AESXPPCLVOAEAEVAYEPEEQETFEQPPVQOQXGSEHIDHHIQGGLSX 482
Db 557 PEEVPAPKSPINXAA-----PIDQ---YDFPFAVABN----- 589
QY 483 QGLCARALYDQADDXEISDPDENLITGIEVXXEGWRRGYPGGHFGKXPANTYVELI 540
Db 590 ----AMALWDYQADDTISFDPDDIITDIDQVDSGWMKGAFSGRVGLFPANTYVELI 643

RESULT 12
US-09-879-957-14
Sequence 14, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOPEMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penile & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mierock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
MOLECULE TYPE: peptide
TOPOLOGY: unknown
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-879-957-14

Query Match 5.2%; Score 208; DB 9; Length 546;
Best Local Similarity 35.6%; Pred. No. 1.3e-09;
Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;
QY 386 SPQPKLSPFLQXLOTPETHFGREPAAXSRPADLPASEPAPSPPCVQAEAEVY 445
Db 407 SQPFLDPP--SSIVEDAPFKAPSPYRGSEF-----EPYS-----IEAGIP 450
QY 446 EEPKQETFEQPPVQOQXGSEHIDHHIQGGLSXGGLCARALYDQADDX 499

Db 451 EAGSQGLTYTSEPVTE-----TTEAPGHVQAEEDYIDGYESPLGITATALYDQAGDD 505
QY 500 EISFDPENLITGIEVXXEGWRRGYPGGHFGKXPANTYVEL 539
Db 506 EISFDPDDITITNEMIDDGWRGV-CKGRYGLFPANTYVEL 544

RESULT 13
US-09-925-299-1277
Sequence 1277, Application US/09925299
Patent No. US2002005627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1277
LENGTH: 40
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (33)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1277

Query Match 4.8%; Score 192; DB 9; Length 40;
Best Local Similarity 90.0%; Pred. No. 1.7e-09;
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 234 WDXTVVROYSFLLMKCCCHGNPAQYERNRRXXHLVYVLG 273
Db 1 WYTVVROYSFLLMKCCCHGNPAQYERNRRXXHLVYVLG 40

RESULT 14
US-09-925-299-1277
Sequence 1277, Application US/09925299
Patent No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1277
LENGTH: 40
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (33)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1277

Query Match 4.8%; Score 192; DB 10; Length 40;
Best Local Similarity 90.0%; Pred. No. 1.7e-09;
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 234 WDXTVVROYSFLLMKCCCHGNPAQYERNRRXXHLVYVLG 273

DB 1 WYTVVROVSTLLMCCCHGNPAQYERRRFXHLVYLG 40

RESULT 15
US-10-369-493-4212
; Sequence 4212, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4212
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(769)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-4212

Query Match 4.8%; Score 190; DB 15; Length 769;
Best Local Similarity 24.9%; Pred. No. 7,4e-08;
Matches 97; Conservative 38; Mismatches 164; Indels 90; Gaps 19;

QY 314 PXXRLSPGPRPLHPRXGRARAP-----PRAGTVRCXVPFVLGNGA 358
DB 414 PEAPASDAG--ISALREFKDTAFAVARTGSRASPSVSPGATGEGHTPPPLTGS 470
QY 359 RPE--RRPYKRPSTARSMTWRTXSISPPQPK-LSPFLQXQLTOPETHGREPA- 413
DB 471 RPSGGFALPGL-PSRP-----SVADDEGNDEROLEEVRALPPADQHEEPSS 518
QY 414 -AXXSRL--PADLPAEPPASXPCVQAEAEVVE--PXQETFEQPLVQOQXGS 468
DB 519 PARIAVEIASAVPEVEPPAGNLPPRPVPSQDLPKEEDLPEEDT--HDPRAALNVAS 575
QY 469 EHIDHITQGGLSXQ-----GLCARALYDQAADXEISFPDENLITGIEVXXEGWVG 522
DB 576 ESL-----GGQVEVQAGSDGGRALVQYDYEKADNEIDLQEGDYVTNIQWVDDMMWG 630
QY 523 YGPDGHTGKXKPAVVEILEXGHIILXFPSTQTLVYCKGRKPKXXHSALFQEXDPQXGX 582
DB 631 TNAQGESGLPPSNVVEVEDEPPAAPSAPT--VAATQAPVPSAQAEAV----- 680
QY 583 LRAPSGIADSACHFCKX-----NG--LVIPTPSCIPRPXX-----TAWLLPLT 624
DB 681 --APATPADQVAGTATAQFDYEAEDNGKXCLEAPHISAKRPPSSDXRIGTELSFP-E 737
QY 625 GYXAKPCXWPSPEWPLPSCGEGSXGASG 653
DB 738 GATITGDEFDDDMWF-----GHYSASG 761

Search completed: April 27, 2004, 11:05:13
Job time : 41.7311 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: April 27, 2004, 10:46:49 ; Search time 17.097 Seconds
(without alignments)
4461.589 Million cell updates/sec

Title: US-10-028-952a-9

Perfect score: 3989

Sequence: 1 HEIPVPTVYPAKQXXERA.....DFRKKKKKKKKLEKWTSSX 793

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	234	5.9	486	2	S07633 hematoopoietic line
2	228.5	5.7	486	2	LckBp1 protein - m
3	226.5	5.7	643	2	T23453 hypothetical prote
4	208	5.2	546	2	I48899 cortactin - mouse
5	204.5	5.1	563	2	A41530 src substrate p80/
6	202	5.1	550	2	A48063 mammary tumor/sgna
7	151.5	3.8	1111	1	A3284 myosin heavy chain
8	148.5	3.7	390	2	G01936 Abl binding protei
9	143	3.6	315	2	A56561 35k proline-rich p
10	143	3.6	592	1	LIBY actin-binding prot
11	141	3.5	290	2	T42526 hypothetical prote
12	134.5	3.4	617	2	S42719 actin-binding prot
13	132	3.3	261	2	S46234 synapain Ia - h
14	131	3.3	705	2	A35363 synapain I eplice
15	126	3.2	6669	2	SS5024 nebulin, skeletal
16	123	3.1	1181	2	T30578 myosin IC - slime
17	121.5	3.0	1107	1	S52517 myosin I heavy cha
18	120	3.0	670	2	S67383 probable signal tr
19	117.5	2.9	540	2	JC4916 signal transducing
20	117.5	2.9	704	2	A30411 synapain Ia - rat
21	116	2.9	1270	2	T09194 adaptor protein in
22	115.5	2.9	1290	2	A31317 1-phosphatidylinos
23	115.5	2.9	1291	2	S00666 1-phosphatidylinos
24	114.5	2.9	1168	1	MMAXIC myosin heavy chain
25	114	2.9	239	2	A46243 epidermal growth f
26	114	2.9	304	2	I58394 c-Crk - mouse
27	114	2.9	1109	2	S53601 myosin-IC (similar
28	113.5	2.8	330	2	UB0376 Grb-2 related adap
29	113.5	2.8	452	2	S46798 hypothetical prote

30	113.5	2.8	1290	2	A36465 1-phosphatidylinos
31	112.5	2.8	548	2	UC4917 signal transducing
32	112	2.8	304	2	A45022 CRK-II human
33	112	2.8	1099	2	A59300 myosin-I δ - mouse
34	111.5	2.8	269	2	UC4899 proline rich prote
35	111.5	2.8	1099	1	S31926 myosin IB heavy ch
36	111.5	2.8	1215	2	T32734 myosin-IA - Acanth
37	110.5	2.8	825	2	T23612 hypothetical prote
38	110	2.8	940	2	T00056 hypothetical prote
39	110	2.8	1006	2	T00050 hypothetical prote
40	109	2.7	1460	1	EDB81F immediate-early pr
41	108.5	2.7	367	2	S50621 SSu1 protein - ye
42	108.5	2.7	629	2	T14776 hypothetical prote
43	107.5	2.7	209	2	S44650 f42n10.3 protein -
44	106	2.7	303	2	S58352 SH2/SH3 adaptor pr
45	105	2.6	379	2	S31719 proline-rich prote

ALIGNMENTS

RESULT 1

S07633 hematoopoietic lineage cell-specific protein Hs1 - human
C/Species: Homo sapiens (man)
C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #ext_change 05-Nov-1999
C/Accession: S07633; A47478; B47478; C47478; D47478; E47478
R/Kitamura, D.; Kaneko, H.; Miyagoe, Y.; Ariyasu, T.; Watanabe, T.
Nucleic Acids Res. 17, 9367-9379, 1989
A/Title: Isolation and characterization of a novel human gene expressed specifically in
A/Reference number: S07633; M01D:90067924; P01D:2587259
A/Accession: S07633
A/Molecule type: mRNA
A/Residues: 1-486 <KIT>
A/Cross-references: EMBL:X16663; NID:932054; PIDN:CAA34651.1; PID:932055
R/Yamanashi, Y.; Okada, M.; Samba, T.; Yamori, T.; Umemori, H.; Tsunashima, S.; Toyoshita
Proc. Natl. Acad. Sci. U.S.A. 90, 3631-3635, 1993
A/Title: Identification of Hs1 protein as a major substrate of protein-tyrosine kinase
A/Reference number: A47478; M01D:93234551; P01D:7682714
A/Accession: A47478
A/Status: preliminary
A/Molecule type: protein
A/Residues: 4-19, 'XXX', '23-26', 'X', '79-93', 'X', '95', 'X', '134-146', '208-223', '274-284', 'X', '286', 'X', '2
A/Experimental source: Daudi, B-lymphoblastoid cells
A/Note: sequence modified after extraction from NCBI backbone
C/Superfamily: SH3 homology
F/435-482/Domain: SH3 homology <SH3>

Query Match 5.9%; Score 234; DB 2; Length 486;

Best local Similarity 26.8%; Pred. No. 1.8e-12;

Matches 84; Conservative 25; Mismatches 97; Indels 108; Gaps 12;

QY	275	GANG--AKLXVGLXNASKSESRFGXITRORGSVGLGXKXKXSPAGRP-----	325
DB	234	GARGLAKRKSVAEKRKEEERKAQVARROQERRA-----VTKSSPAPQVIMSEB	288
QY	326	--PLHAPKXGRARAPRAGTGVRCVFPVLGAGAPERRPXPRTARSWGTWXTXS	383
DB	289	AVAPAPPKTKSSAMP-----VGTPTSE--SEPTATREHNV-----	325
QY	384	ISSPQPKLRSPPLQXOLTOPETHFGREPAAXXSRRADLPABEPAPSPAPCL--VOAE	440
DB	326	-----PLLPTRQTLPEDN-----EEP-PALFPRTLGLQVE	355
QY	441	EEAVVEEPKQERTFYQPPVVOQQXXGSEHIDHIO-----	476
DB	356	EEVVTAEAPPP--EPREPPNDYDEVEMDNHEBDEPGEVEVLPEBDSFSSALA	413
QY	477	-----GGGSKXQGLCARALDYQADDXKISFDPENLTGIVXXEGWRRGCGPG	527
DB	414	GSSGCPAGAGAVALGISVALVDYQGGSDLSFDPDVTIDIMVDEGWRRG--RCGG	472
QY	528	HFGMPKANTVELLE	541

Db 473 HFGLFPANVYKLE 486

RESULT 2

LckBp1 protein - mouse
N:Alternate names: HSI
C:Species: Mus musculus (house mouse)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
A:Accession: I49760; S57244; S52474
R:Kikemura, D.; Kaneko, H.; Tanisuchi, I.; Akagi, K.; Yamamura, K.; Watanabe, T.
Biochem. Biophys. Res. Commun. 208, 1137-1146, 1995
A>Title: Molecular cloning and characterization of mouse HSI.
A:Reference number: I49760; MUID:95217198; PMID:7535527
A:Accession: I49760
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-486 <RES>
A:Cross-references: GB:ID42120; NID:9806521; PIDN:BA07701.1; PID:9806522
R:Takemoto, Y.; Furuta, M.; Li, X.K.; Strong-Sparks, W.O.; Hashimoto, Y.
EMBO J. 14, 3403-3414, 1995
A>Title: LckBp1, a proline-rich protein expressed in haematopoietic lineage cells, direct
A:Reference number: S57244; MUID:95354658; PMID:7628441
A:Accession: S57244
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-486 <RES>
A:Cross-references: EMBL:X84797; NID:9683480; PIDN:CA59265.1; PID:9683481
C:Superfamily: SH3 homology
F:436-483/Domain: SH3 homology <SH3>

Query Match 5.7%; Score 228.5; DB 2; Length 486;
Best Local Similarity 29.5%; Pred. No. 5.3e-12;
Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;
275 GANG-AKLXSVGLINAKSKESRPGXITRQR--RCASVGLGXPPXRLSPPAGRP--PLH 328
234 GAGGLAKFESLAEBKREBEKKAQOMARQOCERKAVVMREVRQCPMPYEEPAAPQ 293
QY 323 APXKRGARAPPAGGCVFVGLGNARPERRRXXKPSSTARWGTWRTXGISSPQ 388
Db 294 LPKKISSEWVP-----AESHLPESSQPVSSRREY-----DV 325
QY 389 PG-KLRSPFQXQLOTPETFGREPAAXSR--PRADLPBEPAPSPCLVQAEBAVYE 446
Db 326 PSLPTQSPLOHLEHNE-----EPFALPRTREGQVVEEPYEAAPL--EPEPEPDYE 379
QY 447 EPEKQETFEQPLVQOQXXGSEHIDHNI-----QGQGLSXQGLCARALYD 492
Db 380 PEFETPEPDYEDVGLRQDEADAGDYEDVLEPEPTSLSYQAGPSAGAGISATAYD 439
QY 493 YQADADKEISFPDENLITGIEVXXEGWMRGYCPDGHFGKXPANYVELI 540
Db 440 YQGGSDLSFDPDITITDIEMDDEGMKRGQ--CRGHGLFPANVYKLE 486

RESULT 3

hypothetical protein K08E3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A:Accession: T23453
R:McMurray, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19743
A:Accession: T23453
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-643 <ML>
A:Cross-references: EMBL:Z81568; PIDN:CA804592.1; GSPDB:GN00021; CESP:K08E3.4
C:Experimental source: clone K08E3
C:Genetics:

A:Gene: CESP:K08E3.4
A:Map position: 3
A:Introns: 47/1; 84/3; 185/2; 255/2; 455/1

Query Match 5.7%; Score 226.5; DB 2; Length 643;
Best Local Similarity 27.5%; Pred. No. 1.1e-11;
Matches 82; Conservative 20; Mismatches 75; Indels 121; Gaps 13;

QY 319 SPAGRPPLHAPKXRGARAPPAGGCVFVGLGNARPERRRXXR-----PS----- 369
Db 391 SKPSG-FVLFKPVQVNSPKKMPVGT-----TSPREVNLTEDDESTYT 436
QY 370 -----RTARSGTWRX-----TXSISPOGKLRSP--FL 397
Db 437 PKLIAPEPEMYKPEPMKPAVSYDAVEBPAPAPPTLAPPLVAPPPBPSTAPSHY 496
QY 398 QXQLOTPETFGREPA--XXSR-----RADLPA----- 425
Db 497 ASQYDAPPVHESEFEPAPVPVSAFSAVSAQYDAPPEDIDSHSSQLPAHIAQYDMPVW 556
QY 426 -EFP--APSKPCLVQAEBAVYEPEPEKQETFEQPLVQOQXXGSEHIDHNIQOGLSX 482
Db 557 PEEVFPAPKSPFKVAA-----PIDQ-----YDFPVAEAQN----- 589
QY 483 QGLCARALYDQADDXEISFPDENLITGIEVXXEGWMRGYCPDGHFGKXPANYVELI 540
Db 590 -----AMALMDYQADDTIEISFPDITITDIDQVDSGWMKGRAPSGVGLFPANVYKLI 643

RESULT 4

148899
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
A:Accession: I48899; A49514
R:Milatarese, M.R.; Manton-Henderson, J.; Wu, H.; Parsons, J.T.; Bender, T.P.
Oncogene 9, 1989-1997, 1994
A>Title: The protein tyrosine kinase substrate contractin is differentially expressed in
A:Reference number: I48899; MUID:94268839; PMID:7516062
A:Accession: I48899
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-546 <RES>
A:Cross-references: EMBL:U03184; NID:9414990; PIDN:AA19689.1; PID:9509495
R:Zhan, X.; Hu, X.; Hampton, B.; Burgess, W.H.; Friesel, R.; Maciag, T.
J. Biol. Chem. 268, 24427-24431, 1993
A>Title: Murine contractin is phosphorylated in response to fibroblast growth factor-1
A:Reference number: A49514; MUID:94043284; PMID:7693700
A:Accession: A49514
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-8, 'R', 10-297, 'S', 299-546 <ZHA>
A:Experimental source: BALB/c 3T3 cells
A>Note: sequence extracted from NCBI backbone (NCBI:P.139492)
C:Superfamily: SH3 homology
F:495-542/Domain: SH3 homology <SH3>

Query Match 5.2%; Score 208; DB 2; Length 546;
Best Local Similarity 35.6%; Pred. No. 3.6e-10;
Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;

QY 386 SPQFKLSPFLQLOTPETFGREPAAXSRPRADLPBEPAPSPCLVQAEBAVY 445
Db 407 SPQFIEDP--SSPIYEDAPFAFABSRYSSB-----EPEYS-----TEAGIP 450
QY 446 EPEKQETFEQPLVQOQXXGSEHIDHNIQOQ-----LSXQGLCARALYDQADDX 499
Db 451 EAGSQQGLTYSEPEYR-----TEAPGHVQABDITDGHSDGCTIRALYDQAGDD 505
QY 500 EISFPDENLITGIEVXXEGWMRGYCPDGHFGKXPANYVEL 539
Db 506 EISFPDITINIMIDGMWGRV--CKGRVGLFPANVYEL 544

RESULT 5

A41530
src substrate p80/85 - chicken
N:Alternate names: kinase-related transforming protein substrate
C:Species: Gallus gallus (chicken)
C>Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 21-Jul-2000
C/Accession: A41530
R/Mu. H.; Reynolds, A.B.; Kanner, S.B.; Vines, R.R.; Parsons, J.T.
Mol. Cell. Biol. 11, 5113-5124, 1991
A/Title: Identification and characterization of a novel cytoskeleton-associated pp60(src
A/Reference number: A41530; MUID:92017786; PMID:1922035
A/Accession: A41530
A/Molecule type: mRNA
A/Residues: 1-563 <MUA>
A/Cross-references: GB:M73705; NID:g212588; PIDN:AAA49031.1; PID:g212589
C/Superfamily: SH3 homology
C/Keywords: phosphoprotein; tandem repeat
F:92-128/Domain: src substrate repeat <SR1>
F:129-165/Domain: src substrate repeat <SR2>
F:166-202/Domain: src substrate repeat <SR3>
F:203-239/Domain: src substrate repeat <SR4>
F:240-276/Domain: src substrate repeat <SR5>
F:277-313/Domain: src substrate repeat <SR6>
F:314-334/Domain: src substrate repeat #status atypical <SR7>
F:512-559/Domain: SH3 homology <SH3>

Query Match 5.1%; Score 204.5; DB 2; Length 563;
Best Local Similarity 36.1%; Pred. No. 7.4e-10;
Matches 52; Conservative 18; Mismatches 45; Indels 29; Gaps 5;

QY 424 PAEPAPSP--PCLVQAEBAVEEPEX-----EGETFE-----QPLVQOQXXGS 468
DB 419 PAEPKTPSPVYQAVASVDAESAYKNSSTYSAEHPEBSCYKTTGSDYQAVSQREAYE 478
QY 469 EHIHHTQOGLSKXQ-----GLCARLYVQADDXEISFDPENLITGLEYX 515
DB 479 PETVEVAGADHYQAEENTYDEYENELGITVAILYDQAGDEISFPDDITITENT 538
QY 516 XEGWNRGYGPDGHFGKXPANYVEL 539
DB 539 DDGWMRGV--CKGRYGLFPANYVEL 561

RESULT 6

A48063
mammary tumor/squamous cell carcinoma-associated protein EMS1 - human
C/Species: Homo sapiens (man)
C>Date: 24-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 29-Aug-1997
C/Accession: A48063
R/Schunung, E.; Verhoeven, E.; Litvinov, S.; Michalides, R.J.
Mol. Cell. Biol. 13, 2891-2898, 1993
A/Title: The product of the EMS1 gene, amplified and overexpressed in human carcinomas,
A/Reference number: A48063; MUID:93233650; PMID:8474448
A/Accession: A48063
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-550 <SCH>
A/Note: sequence extracted from NCBI backbone (NCBI:P.129872)
A/Genetics:
A:Gene: GDB:EMS1
A/Cross-references: GDB:202962; OMIM:164765
A/Map position: 11q13-11q13
C/Superfamily: SH3 homology
F:499-546/Domain: SH3 homology <SH3>

Query Match 5.1%; Score 202; DB 2; Length 550;
Best Local Similarity 36.0%; Pred. No. 1.2e-09;
Matches 58; Conservative 17; Mismatches 60; Indels 26; Gaps 6;
QY 386 SPOGKLRSPFLQQLQTPETHFGEPAAXXSRPADLPAEPAPSPPCLVQAEBAVY 445
DB 407 APQPIERKLP-----SSPV-----YEDASF---KALSTGYSPVSGTEPREVYGMENADY 453

QY 446 EEPXED-----ETPIEOPPLVQOQXXGSEHIDHIOGGLSXQGLCARLYDQADDD 498

DB 454 REASSQOGLAVTEAYE-----SAEPGHYPAEISTYDEYENDIGYTRVALYDQAGD 508

QY 499 XEISFDPENLITGLEYXXEGWNRGYGPDGHFGKXPANYVEL 539
DB 509 DELSPDDITITNEMIDGWMRGV--CKGRYGLFPANYVEL 548

RESULT 7

A33284
myosin heavy chain IB - slime mold (Dictyostelium discoideum)
N:Conting: myosin ATPase (EC 3.6.4.1)
C/Species: Dictyostelium discoideum
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C/Accession: A33284
R/Jung, G.; Saxe III, C.L.; Kimmel, A.R.; Hammer III, J.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 6186-6190, 1989
A/Title: Dictyostelium discoideum contains a gene encoding a myosin I heavy chain.
A/Reference number: A33284; MUID:89345628; PMID:2762320
A/Accession: A33284
A/Molecule type: DNA
A/Residues: 1-1111 <JUN>
A/Cross-references: GB:M26037; NID:g167838; PIDN:AAA33229.1; PID:g167839
C/Genetics:
A/Intons: 1/3; 39/3
C/Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homol
C/Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein;
F:12-678/Domain: myosin motor domain homology <MOT>
F:102-109/Region: nucleotide-binding motif A (P-loop)
F:566-587/Region: actin binding #status predicted
F:712-1111/Domain: carboxyl-terminal <CTD>
F:712-901/Region: basic
F:922-1058/Region: alanine/glutamine/glycine/proline-rich
F:1060-1108/Domain: SH3 homology <SH3>
F:108/Binding site: ATP (Lys) #status predicted
F:332/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.8%; Score 151.5; DB 1; Length 1111;
Best Local Similarity 21.7%; Pred. No. 5.9e-05;
Matches 57; Conservative 21; Mismatches 86; Indels 99; Gaps 5;

QY 291 SKSESPPGXTIRQAGASVGLGXPPXRLSPRPPLNAPRGARAPPRAGTGRXCV 350
DB 931 SQSSKPA---PQSAKRGKMPQG---AGQPPQPPQOQRPMPQPPQGGARPP 981
QY 351 PVLNGARPERRPXRRPRTARSWGTXYSISSPPQGRKURSPFLQQLQTPETHRGR 410
DB 982 QPQGGGARP-----MGAPQGGAPQOQAGRQLPQPGGA 1018
QY 411 E-----PAAXSRPADLPAEPAPSPPCLVQAEBAVVEEPEXQETFE 455
DB 1019 PGGGAPMGKAPGGGPPAGGRPLPYAKRPQSRP----- 1056
QY 456 EQPPLVQOQXXGSEHIDHITQOGLSXQGLCARLYDQADDXEISFDPENLITGLEYX 515
DB 1057 -----TAKALYDASSTDELSEFGEDIIPIVQGD 1086
QY 516 XEGWNRGYGPDGHFGKXPANYVE 538
DB 1087 NGQWQTEGLKSGQKGAAPTNYLD 1109

RESULT 8

G01936
Abi binding protein 3 - human
C/Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C/Accession: G01936
R/Ren, R.
submitted to the EMBL Data Library, July 1995
A/Reference number: G08875

C/Species: Schizosaccharomyces pombe
 C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T42526
 R/Yoshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
 DNA Res. 4, 363-369, 1997
 A/Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
 A/Reference number: Z17323; MUID:96162722; PMID:9501991
 A/Accession: T42526
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-290 <YOS>
 A/Cross-references: EMBL:D89164; NID:g1749535; PIDN:BA13826.1; PID:g1749536
 A/Experimental source: strain PR745

Query Match 3.5%; Score 141; DB 2; Length 290;
 Best Local Similarity 27.7%; Pred. No. 0.00012;
 Matches 52; Conservative 25; Mismatches 59; Indels 52; Gaps 11;

QY 387 POP-GKLRSPFLQXQLTQPETHFGREPAAXXSRRPADPA-EEBAPSPCLVQAEBAV 444
 DB 121 PEPAGQLEPPVPPPPPHDET---QEPVYG---GDVKAETHTQPTKTPAIV-----I 166
 QY 445 YEEPYOETFEYQEPPLVQ-QQXXGSEHID-----HHIQ-QGL----- 480
 DB 167 YDYSPEENETL---LVENEQLQLEFVDGWLGENSKGQQLFPSNYVEITGPNETAN 223
 QY 461 -----SXQGLCARALYDQADDXISFDPENLIGIEVXXEGWGRGYPGPHGGMX 532
 DB 224 NPPAPQAGGPGKSVKALYDQADDXISFDPENLIGIEVXXEGWGRGYPGPHGGMX 282
 QY 533 PANYVELI 540
 DB 283 PSNYVEEI 290

RESULT 12

S42719
 actin-binding protein ABPI - Saccharomyces exiguus
 C/Species: Saccharomyces exiguus
 C/Date: 07-Sep-1994 #sequence_revision 19-Apr-1996 #text_change 13-Aug-1999
 C/Accession: S42719
 R/Lange, U.; Steiner, S.; Grolig, F.; Wagner, G.; Philippesen, P.
 Biochim. Biophys. Acta 1217, 214-218, 1994
 A/Title: Cloning and sequencing of a gene coding for an actin binding protein of Sacchar.
 A/Reference number: S42719; MUID:94154001; PMID:8110838
 A/Accession: S42719
 A/Molecule type: DNA
 A/Residues: 1-617 <LAN>
 A/Cross-references: EMBL:X73977; NID:g433513; PIDN:CAA52156.1; PID:g433514
 C/Superfamily: actin-binding protein ABPI; SH3 homology
 C/Keywords: actin binding; cytoskeleton
 P.564-613/Domain: SH3 homology <SH3>

Query Match 3.4%; Score 134.5; DB 2; Length 617;
 Best Local Similarity 25.0%; Pred. No. 0.00095;
 Matches 52; Conservative 21; Mismatches 82; Indels 53; Gaps 8;

QY 385 SSPQPKLRSPFLQXQLTQ-PETH-----FGREPAAXXSRR-PADLPAPAE 426
 DB 408 SKPATQEVKKPPTPENIGQLPQKMTETPEHEEDNDGDEDEDEPPKRNIPPPVMPAR 467
 QY 427 EAPSP--KPPCLVQAE--EEAVYEEPYOETFEYQEP-----PLVQOQXXGSEHID 472
 DB 466 ESAPQQLPFPNTEBPVEEGSEEEEBEESAPSLPSRNAAPPEPEQOEEBEE 527
 QY 472 HHIOGGLSXQGLC-----ARALYDQADDXISFDPENLITGI 512
 DB 528 ESAPAPSLPSRGSVPPEPQRAVEPEBPAAEAAPWATAEDYDAGDNELTFAENDKIINI 587
 QY 513 EYXEGWGRG-XGPGHGRMPANYVEL 539
 DB 588 EFVDDDWLIGLETTGQKGLFPSNYVEL 615

RESULT 13

S68234
 Laep-1 protein - human
 N/Alternate names: LIM and SH3 protein
 C/Species: Homo sapiens (man)
 C/Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
 C/Accession: S68234; S72556
 R/Tomasetto, C.; Moog-Lutz, C.; Regnier, C.H.; Schreiber, V.; Basset, P.; Rio, M.C.
 FEBS Lett. 373, 245-249, 1995
 A/Title: Laep-1 (MLN 50) defines a new LIM protein subfamily characterized by the assoc
 A/Reference number: S68234; MUID:96033982; PMID:7589475
 A/Accession: S68234
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-261 <TOM>
 A/Cross-references: EMBL:X82456
 R/Tomasetto, C.
 submitted to the EMBL Data Library, November 1994
 A/Reference number: S72556
 A/Accession: S72556
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-84, 'C', '86-90', 'G', '92-219', 'D', '221-261' <T02>
 A/Cross-references: EMBL:X82456; NID:g951273; PID:g951274
 C/Superfamily: LIM metal-binding repeat homology; SH3 homology
 P.5-56/Domain: LIM metal-binding repeat homology <LIM>
 F.209-258/Domain: SH3 homology <SH3>

Query Match 3.3%; Score 132; DB 2; Length 261;
 Best Local Similarity 29.0%; Pred. No. 0.00064;
 Matches 42; Conservative 17; Mismatches 48; Indels 38; Gaps 7;

QY 409 GREPAAXSRP-----RADPAPERP--APSPCLVQAEBAV-----YEEPYOETFEY 456
 DB 138 GMEPERDSDGSGSYRPLRQQLPHITPSAPYQPPQCPVAQSYGKKEPA----- 190
 QY 457 QPPLVQOQXXGSEHIDHHIQGGLSXQGLCARALYDQADDXISFDPENLITGIEVXX 516
 DB 191 APVSIQRSAPAGC-----GGKRYAVYDYSAADDEVSFGQDGTIVVQQLD 236
 QY 517 EGWGRGYP---DGHGKXPANYVE 538
 DB 237 DGMV--YGVETRTGTGMPDNIVE 259

RESULT 14

A35363
 synapsin I splice form a - human
 C/Species: Homo sapiens (man)
 C/Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 01-Dec-2000
 C/Accession: A35363; B35363; A35805
 R/Suehiof, T.C.
 J. Biol. Chem. 265, 7849-7852, 1990
 A/Title: The structure of the human synapsin I gene and protein.
 A/Reference number: A35363; MUID:90243651; PMID:2110562

A/Accession: A35363
 A/Molecule type: DNA
 A/Residues: 1-705 <SUE>
 A/Cross-references: GB:M58371; GB:J05431
 A/Accession: B35363
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-659, 'KASPAQAP', <SU2>
 A/Cross-references: GB:M58378; GB:J05431
 R/Suehiof, T.C.; Hoeck, C.; Oschwald, R.; Kilian, M.W.
 J. Biol. Chem. 265, 14932-14937, 1990
 A/Title: The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and CAAT-less
 A/Reference number: A35805; MUID:90368667; PMID:2118519
 A/Accession: A35805
 A/Molecule type: DNA
 A/Residues: 1-125 <SNU>
 A/Cross-references: GB:M55301; NID:g338655; PIDN:AAA60608.1; PID:g553654; GB:J05630

C/Genetics:
 A/Gene: GDB:SYN1
 A/Cross-references: GDB:119606; OMIM:313440
 A/Map position: Xp11.23-Xp11.23
 C/Keywords: actin binding; alternative splicing; phosphoprotein

Query Match 3.3%; Score 131; DB 2; Length 705;
 Best Local Similarity 24.7%; Pred. No. 0.0022;
 Matches 76; Conservative 22; Mismatches 126; Indels 84; Gaps 13;

QY 209 VGSSTYXXAEFTKRWIVHPICGNXSWDXTVVRQVSTFLIMCCCHGPNQAYERNRXXHL 268
 DB 388 VGSSTYXXAEFTKRWIVHPICGNXSWDXTVVRQVSTFLIMCCCHGPNQAYERNRXXHL 428
 QY 269 VVVLXGANGAKLXSVGLKIMASKSRSRPGXTTRQRRGASVGLKXP----- 314
 DB 429 ---GRGSHGQRTSPGALPLGRQTSQQAPGPPAQGPPOGPPGPGPQRRGPPLOQR 484
 QY 315 -----XXRLSPRAGR-----PLHAPRXGRARAPPAGTGVRCXVFPVLGNGAR 359
 DB 485 PPGGQGHLSGLPPAPSPPLPQRLPSPFTSAPQQAAPPPQGGQ-RGRVVAAGPGAP 543
 QY 360 PERRPXXRPSRTARSNGTWTXT--XSISPPQPKLR--PFLQXULTQPTHTFG-----R 410
 DB 544 PAARPPAPSP--PQRCAGPPQATRTQTSVSGPAPKASGAPPGGQRRGPPQPKPPGAPPTR 602
 QY 411 EPAAXXSRRPADLP-AEPPAPSP-----PCLVQAEBAVYEEPEQETFEQPPPLVQ 462
 DB 603 QASQAGVPRTGPTTQQPAPSPGPPAGAPKPOLAQKPSQDV-----PPPAT 649
 QY 463 QQXXGSEH 470
 DB 650 AAAGPPH 657

RESULT 15

S55024
 nebulin, skeletal muscle - human
 C/Species: Homo sapiens (man)
 C/Date: 15-Jul-1995 #sequence revision 01-Sep-1995 #text_change 05-Nov-1999
 C/Accession: S55024; S55025; A40847
 R/Label: S. the EMBL Data Library, January 1995
 A/Reference number: S55024
 A/Molecule type: mRNA
 A/Residues: 1-6669 <LAB1>
 A/Cross-references: EMBL:X83957; NID:9806561; PIDN:CA58788.1; PID:9806562
 R/Label: S. Kolmerer, B.
 J. Mol. Biol. 248, 308-315, 1995
 A/Title: The complete primary structure of human nebulin and its correlation to muscle
 A/Reference number: S55025; MUID:95257391; PMID:7739042
 A/Accession: S55025
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 3731-3930:6501-6669 <LAB2>
 A/Cross-references: EMBL:X83957
 R/Label: J. F. Wang, K.
 J. Biol. Chem. 266, 21215-21223, 1991
 A/Title: Cloning, expression, and protein interaction of human nebulin fragments compose
 A/Reference number: A40847; MUID:92042000; PMID:1682316
 A/Accession: A40847
 A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A/Molecule type: mRNA
 A/Residues: 6263-6271, 'M', 6273-6329 <LIN>
 C/Genetics:
 A/Gene: GDB:NEB
 A/Cross-references: GDB:120224; OMIM:161650
 A/Map position: 2q24.1-2q24.2
 C/Superfamily: human skeletal muscle nebulin; SH3 homology
 C/Keywords: alternative splicing; skeletal muscle; tandem repeat
 F/78-6531/Region: 35-residue repeats
 F/6617-6666/Domain: SH3 homology <SH3>

Query Match 3.2%; Score 126; DB 2; Length 6669;
 Best Local Similarity 27.0%; Pred. No. 0.061;
 Matches 47; Conservative 13; Mismatches 72; Indels 42; Gaps 5;
 QY 402 TQETTFGRPPAAXXSRRP-----ADLPABEPAPSPCLVQAE----- 442
 DB 6503 TNPQSVDDYDPADNITQSRSLHMIVQARRSRQSRASALSVSGEEKSEHSEAPDHH 6562
 QY 443 -----AVYEPKQETFYHQPPLVQOQXGSEHIDHITQGGLSXQGLCARA 489
 DB 6563 LSTYSDGVAVSTAYKHATTELPOQRSSSVATQOTTSSIPSH-----PSTAGKIFRA 6617
 QY 490 LYDYQADDDXISFPENLITGIEVXXEGMWRGYGP---DGHGKMPANYVELI 540
 DB 6618 MYDMADADAVSFKQGDALINQALDEGMN--YGTVQRTGRTGMLPANYVELI 6669

Search completed: April 27, 2004, 10:56:01
 Job time : 19.097 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:25:14 ; Search time 14.5828 Seconds
(without alignments)
2831.539 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989
Sequence: 1 HEIPVPTVPPAKPQXXERA.....DFRKKKKKKLEKLEMTSSX 793

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	234	5.9	486	1	HS1_HUMAN P14317 homo sapien
2	228.5	5.7	486	1	HS1_MOUSE Q69710 mus musculu
3	208	5.2	546	1	SRCS_MOUSE Q60598 mus musculu
4	204.5	5.1	563	1	SRCS_CHICK Q01406 gallus gall
5	202	5.1	550	1	SRCS_HUMAN Q14247 homo sapien
6	151.5	3.8	1111	1	MYCB_DICDI P34092 dictyosteli
7	148	3.7	263	1	LSL1_MOUSE Q61792 mus musculu
8	143.5	3.6	857	1	APB1_SCHPO Q95768 schizosacch
9	143	3.6	592	1	APB1_YEAST P15891 saccharomyc
10	141	3.5	261	1	LSL1_HUMAN Q14847 homo sapien
11	134.5	3.4	617	1	ABP1_SACEX P38479 saccharomyc
12	131	3.3	705	1	SYN1_HUMAN P17600 homo sapien
13	129.5	3.2	1098	1	MYIF_HUMAN Q00160 homo sapien
14	126	3.2	6659	1	NEB1_HUMAN P20928 homo sapien
15	124.5	3.1	1217	1	ITN1_RAT Q94074 mus musculu
16	124	3.1	1714	1	ITN1_MOUSE Q62419 mus musculu
17	123	3.1	368	1	SH31_MOUSE P42522 dictyosteli
18	123	3.1	1181	1	MYSC_DICDI Q88935 mus musculu
19	122.5	3.1	670	1	SYN1_MOUSE Q63356 rattus norv
20	121.5	3.0	1107	1	MYIE_RAT Q99961 homo sapien
21	121	3.0	368	1	SH31_HUMAN Q10199 schizosacch
22	120	3.0	670	1	YBY2_SCHPO Q76041 homo sapien
23	120	3.0	1014	1	NEB1_HUMAN Q15811 homo sapien
24	119	3.0	1721	1	ITN1_HUMAN Q09951 rattus norv
25	117.5	2.9	704	1	SYN1_RAT Q61644 mus musculu
26	116.5	2.9	441	1	PAC1_MOUSE Q94075 r protein k
27	116.5	2.9	441	1	PAC1_RAT Q99111 homo sapien
28	116.5	2.9	444	1	PAC1_HUMAN Q15735 homo sapien
29	116.5	2.9	1006	1	P5PA_HUMAN Q9448 rattus norv
30	116.5	2.9	2167	1	SHK1_RAT Q63768 rattus norv
31	116	2.9	304	1	CRK_RAT Q9486 homo sapien
32	116	2.9	424	1	PAC3_HUMAN Q42287 xenopus lae
33	116	2.9	1270	1	ITN1_XENLA

34	115.5	2.9	1290	1	P1G1_RAT P10686 rattus norv
35	115.5	2.9	1291	1	P1G1_BOVIN P08487 bos taurus
36	114.5	2.9	486	1	PAC2_MOUSE Q94917 rattus norv
37	114.5	2.9	488	1	PAC2_RAT Q94917 rattus norv
38	114.5	2.9	1168	1	MYSC_ACACA Q64010 mus musculu
39	114	2.9	304	1	CRK_MOUSE Q05433 avian sarco
40	114	2.9	440	1	GAGC_AYISC Q12365 homo sapien
41	114	2.9	1109	1	MYIE_HUMAN Q75791 h grb2-rela
42	113.5	2.8	330	1	GRP2_YEAST P38753 saccharomyc
43	113.5	2.8	452	1	YH2_YEAST P19174 homo sapien
44	113.5	2.8	1290	1	P1G1_HUMAN Q10571 homo sapien
45	113.5	2.8	1319	1	MN1_HUMAN

ALIGNMENTS

RESULT 1	HS1_HUMAN	STANDARD;	PRT;	486 AA.
AC P14317				
DT 01-JAN-1990 (Rel. 13, Created)				
DT 01-JAN-1990 (Rel. 13, Last sequence update)				
DT 15-MAR-2004 (Rel. 43, Last annotation update)				
DE Hematopoietic lineage cell specific protein (Hematopoietic cell-specific lym substrate 1) (ICKBP1).				
GN HCKSL OR HS1.				
OS Homo sapiens (Human).				
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OC NCBI_Taxid=9606;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=90067934; PubMed=2587259;				
RA Kitamura D., Kaneo H., Miyagoe Y., Ariyasu T., Watanabe T.,				
RT "Isolation and characterization of a novel human gene expressed specifically in the cells of hematopoietic lineage."				
RL Nucleic Acids Res. 17:9367-9379(1989).				
RN [2]				
RP SEQUENCE FROM N.A.				
RC TTSQB=Lympn;				
RX MEDLINE=22388257; PubMed=12477932;				
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,				
RA Altschul S.F., Zeeberg B., Buettner R.H., Schaefer C.F., Bhat N.K.,				
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA Diatchenko L., Marstina K., Farmer A.A., Rubin G.M., Hong L.,				
RA Spletton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.J.,				
RA Raha S.S., Loggellano N.A., Peters G.U., Abramson R.D., Mulaly S.J.,				
RA Besak S.A., McKernan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,				
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA Fahy J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,				
RA Witting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA Blakesley R.W., Groomwood J.W., Green E.D., Dickson M.C.,				
RA Rodriguez A.C., Richmond J., Schmutz J., Myers R.M.,				
RA Butcherfield Y.S.N., Krzyzanski M.I., Skalska U., Smalins D.B.,				
RA Scheraga A., Scheraga J.E., Jones S.J.M., Marra M.A.,				
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."				
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN [3]				
RP SEQUENCE OF 97-108, 193-201 AND 240-248.				
RX MEDLINE=96311348; PubMed=8713105;				
RA Egerton M., Moritz R.B., Drucker B., Kelso A., Simpson R.J.,				
RT "Identification of the 70KD heat shock cognate protein (Hsc70) and alpha-actinin-1 as novel phosphotyrosine-containing proteins in T lymphocytes."				
RL Biochem. Biophys. Res. Commun. 224:666-674(1996).				
RN [4]				
RP BINDING TO HAX-1 PROTEIN.				
RX MEDLINE=97211841; PubMed=9058608;				

RA Suzuki Y., Demoliere C., Kitamura D., Takeshita H., Deuschle U.,
 RA Watanabe T.;
 RT "HAX-1", a novel intracellular protein, localized on mitochondria,
 RT directly associates with Hs1, a substrate of Src family tyrosine
 RT kinases.";
 RL J. Immunol. 158:2736-2744(1997).
 RN [5]
 RP PHOSPHORYLATION OF TYR-222.
 RX MEDLINE=99167527; PubMed=10066823;
 RA Brunacci A.M., Donella-Deana A., James P., Quadroni M., Contini A.,
 RA Martin O., Pinna L.A.;
 RT "Molecular features underlying the sequential phosphorylation of Hs1
 RT protein and its association with c-Fgr protein-tyrosine kinase.";
 RL J. Biol. Chem. 274:7587-7594(1999).
 CC -1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine
 CC kinase. Plays a role in antigen receptor signaling for both
 CC clonal expansion and deletion in lymphoid cells. Directly
 CC associates with HAX-1, through binding to its C-terminal region.
 CC MAY ALSO BE INVOLVED IN THE REGULATION OF GENE EXPRESSION.
 CC -1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.
 CC -1- TISSUE SPECIFICITY: Expressed only in tissues and cells of
 CC hematopoietic origin.
 CC -1- DEVELOPMENTAL STAGE: Expressed in early stage of myeloid and
 CC erythroid differentiation.
 CC -1- PTM: Phosphorylated by LYN, rapidly after crosslinking of surface
 CC IGM on B cells.
 CC -1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16663; CAA34651.1; -;
 DR EMBL: BC016758; AAH16758.1; -;
 DR FTR: S07633; S07633.
 DR HSSP: P23355; ISEM.
 DR Genew: HGNC:4844; HCLSI.
 DR MIM: 601306; -;
 DR GO: GO:0005665; C:DNA-directed RNA polymerase II, core complex; TAS.
 DR GO: GO:0003700; F:transcription factor activity; TAS.
 DR GO: GO:0007242; F:intracellular signaling cascade; TAS.
 DR GO: GO:0006355; P:regulation of transcription; DNA-dependent; TAS.
 DR InterPro: IPR00134; Hs1/Cortactin.
 DR InterPro: IPR00108; Neu_cyt_fact_2.
 DR Pfam: PF02218; Hs1_rep; 4.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00459; P67PFX.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000066; SH3; 1.
 DR SMART: SMO0326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Repeat: SH3 domain; Phosphorylation.
 KW Repeat: SH3 domain; INVOLVED IN HAX-1 BINDING.
 FT DOMAIN 27 66
 FT REPEAT 81 214 3.5 X 37 AA TANDM REPEATS.
 FT REPEAT 81 116 1.
 FT REPEAT 117 153 2.
 FT REPEAT 154 190 3. (INCOMPLETE).
 FT REPEAT 191 214 4.
 FT DOMAIN 428 486 SH3.
 FT MOD RES 222 222 PHOSPHORYLATION.
 FT MOD RES 241 242 KF -> FK (IN REF. 3).
 FT CONFLICT 241 242
 SQ SEQUENCE 486 AA; 5398 MM; 61AE637157DF5DF2 CRC64;
 Query Match 5.9%; Score 234; DB 1; Length 486;
 Best Local Similarity 26.8%; Freq. No. 4.5e-12;
 Matches 84; Conservative 25; Mismatches 97; Indels 108; Gaps 12;

QY 275 GANG--AKLXSVGLXINASKSESPGKXITRORGASVGLGKPPXRLSPAGRP----- 325
 DB 234 GARGLKAKPESSMAEKREKEERKAQVARROQERRA-----VTKSPSPAPVIMEEP 288
 QY 326 --PLHAPRXGRARAPPAGATGVRCVFPVLGNGARPERPXXRPSTARSQWXTXS 383
 DB 289 AVPAAPLPKTISSRAMP-----VGTPSSP--SEFVTRSEHPV----- 325
 QY 384 ISSPQPKLSPPLQQLQPTFFHGRPAAXSRPADIPAEPPASPPCL---VOAE 440
 DB 326 -----PLPLPRTLPEDN-----SEP-PAUPRTLBGLQVE 355
 QY 441 EEAVERPEXQEPFYEQPPVVOQXXGSEHIDHIQ----- 476
 DB 356 EEPVTEPEPPPP--EPPPPENDYEDVEMDMHDEDEPPGDEVELEPPDSFSSALA 413
 QY 477 -----GQGLSXQGLCARALDYQADDXISTDPENLTIGIEVXXEGWVRGYGPDG 527
 DB 414 GSSGCPAGAGAGAVALGISAVALDYQEGSDLSFDPDVIYDIMEVDEGMWG--RCHG 472
 QY 528 HFGKXPANYVELLE 541
 DB 473 HFGLPANYVALLE 486
 RESULT 2
 Hs1_MOUSE
 ID Hs1_MOUSE STANDARD; PRT; 486 AA.
 AC P45710
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hematopoietic lineage cell specific protein (Hematopoietic cell-
 DE specific LYN substrate 1) (LCKBPI).
 GN HCLSI OR Hs1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95217198; PubMed=7535527;
 RX Kitamura D., Kaneko H., Tanuchi I., Yamamura K.I., Watanabe T.;
 RA "Molecular cloning and characterization of mouse Hs1.";
 RL Biochem. Biophys. Res. Commun. 208:1137-1146(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RX MEDLINE=97098949; PubMed=8943564;
 RA Takemoto Y., Sato M., Furuta M., Hashimoto Y.;
 RT "Distinct binding patterns of Hs1 to the Src SH2 and SH3 domains
 RT reflect possible mechanisms of recruitment and activation of
 RT downstream molecules.";
 RL Int. Immunol. 8:1699-1705(1996).
 CC -1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine
 CC kinase. Plays a role in antigen receptor signaling for both clonal
 CC expansion and deletion in lymphoid cells. Directly associates with
 CC HAX-1, through binding to its C-terminal region. May also be
 CC involved in the regulation of gene expression (By similarity).
 CC -1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.
 CC -1- TISSUE SPECIFICITY: Expressed only in tissues and cells of
 CC hematopoietic origin.
 CC -1- PTM: Phosphorylated by LYN, rapidly after crosslinking of surface
 CC IGM on B cells (By similarity).
 CC -1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D42120; BAA07701.1; -
 DR EMBL; X84797; CAA59265.1; -
 DR PIR; I49760; I49760.
 DR HSSP; P29355; 1SEM.
 DR MGD; MGI:104568; Hc1a1.
 DR GO; GO:0017124; F:SH3-domain binding; IDA.
 DR InterPro; IPR003134; Hsl/Coractin.
 DR InterPro; IPR000108; Neu_cyt_fact_2.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF02218; Hsl_rep; 4.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00459; P67PHOX.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR Repeat; SH3 domain; Phosphorylation.
 KW Repeat; SH3 domain; Phosphorylation.
 FT DOMAIN 27 66 INVOLVED IN HAX-1 BINDING (BY
 FT SIMILARITY).
 FT 27 66
 FT DOMAIN 81 214 3.5 X 37 AA TANDEM REPEATS.
 FT REPEAT 81 116 1.
 FT REPEAT 117 153 2.
 FT REPEAT 154 190 3.
 FT REPEAT 191 214 4 (INCOMPLETE).
 FT DOMAIN 429 486 SH3.
 FT MOD RES 222 222 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 486 AA; 54212 MW; 1ESD05D0333AA0F CRC64;

Query Match 5.7%; Score 228.5; DB 1; Length 486;
 Best Local Similarity 29.5%; Pred. No. 1.3e-11;
 Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GANQ--AKLXSVGLKLNKSKSRSGXITRQR--RGASVGLGXPRXRISSPAGRP--PLH 328
 DB 234 GARGKAKFESLAEEKRKEEEKKQQAARQQQEKAVKRSREVQSSMVEEPADPAQ 293
 QY 329 APRXGRAPAPPAGTGVRCXVPVLGNGARPERRRPXXRPSRTASMGTXISISSPQ 388
 DB 294 LFKKISSEWMP-----AESHLPESSQPVXRSREY-----FV 325
 QY 389 PG-KLRSPFLQXQLQPTETFGREPAAXSR-PRADLAEPPAPSPCLVQAEAEAYE 446
 DB 326 PSLPTROSLQWHLADNE---EPPALPPTPEGLQVEEVEYEADEL-EPEEPDYE 379
 QY 447 EFXEQTFTYEQPPLVQOQXGSEHIDHNI-----CGQGLSXQGLCARALYD 492
 DB 380 PEPETPEPDEVDGELDRQEDAEAGDEVDLEBEDTPSLSYQAGPAGAGGAGISALAYD 439
 QY 493 YQADDXEISFDPENLITGIEVXXEGGWRGYPDGHFGKXPANYVELI 540
 DB 440 YQEGSGDELSFDPDIIITDIEMVDEGWRGQ-CRSHFGFLPANYVELI 486

RESULT 3
 SRC8_MOUSE STANDARD; PRT; 546 AA.
 ID SRC8_MOUSE
 AC Q60598;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Src substrate coractin.
 OS Ems1 OR CTN.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;

RX MEDLINE=94268839; PubMed=7516062;
 RA Mlgrease M.R., Mannion-Henderson J., Wu H., Parsons J.T.,
 RA Bender T.P.;
 RT "The protein tyrosine kinase substrate coractin is differentially
 RT expressed in murine B lymphoid tumors.";
 RL Oncogene 9:1989-1997(1994).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 125-138; 273-289 AND 534-543.
 RX MEDLINE=94043284; PubMed=7693700;
 RA Zhan X., Hu X., Hampton B., Burgess W.H., Friesel R., Maciag T.;
 RT "Marine coractin is phosphorylated in response to fibroblast growth
 RT factor-1 on tyrosine residues late in the G1 phase of the BALB/c 3T3
 RT cell cycle.";
 RL J. Biol. Chem. 268:24427-24431(1993).
 CC -1- FUNCTION: May contribute to the organization of cell structure.
 CC The SH3 motif may function as a binding region to cytoskeleton.
 CC Tyrosine phosphorylation in transformed cells may contribute to
 CC cellular growth regulation and transformation.
 CC -1- SUBUNIT: Interacts with SHANK2 and SHANK3 via its SH3 domain (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Detected in most murine tissues, but not
 CC detected in B lymphocytes or plasma cells.
 CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
 CC -1- SIMILARITY: Contains 1 SH3 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U03184; AAA18689.1; -
 DR PIR; I48899; I48899.
 DR HSSP; P19174; 2HSP.
 DR MGD; MGI:99695; Ctn.
 DR InterPro; IPR001134; Hsl/Coractin.
 DR InterPro; IPR000108; Neu_cyt_fact_2.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF02218; Hsl_rep; 7.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00459; P67PHOX.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR Repeat; SH3 domain; Cytoskeleton.
 KW Phosphorylation; Repeat; SH3 domain; Cytoskeleton.
 FT DOMAIN 83 329 6.5 X 37 AA TANDEM REPEATS.
 FT REPEAT 83 119 1.
 FT REPEAT 120 156 2.
 FT REPEAT 157 193 3.
 FT REPEAT 194 230 4.
 FT REPEAT 231 267 5.
 FT REPEAT 268 304 6.
 FT REPEAT 305 329 7 (INCOMPLETE).
 FT DOMAIN 488 546 SH3.
 FT CONFLICT 9 A -> R (IN REF. 2).
 SQ SEQUENCE 546 AA; 61260 MW; 8F93A026AD1D6D4F CRC64;

Query Match 5.2%; Score 208; DB 1; Length 546;
 Best Local Similarity 35.6%; Pred. No. 7.4e-10;
 Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;

QY 386 SPQGRKLRSPFLQXQLQPTETFGREPAAXSRPRADLPAEPAPSPCLVQAEAEAYE 445
 DB 407 SPQIEDRP--SSPIEDAPPAEPSPYSGSEP-----EPEYS-----IEAGIIP 450
 QY 446 EFXEQTFTYEQPPLVQOQXGSEHIDHNIQGG-----LSYQGLCARALYQADDX 499
 DB 451 EAGSQQGLTYSEPYE-----TTEAPGHYQADDDTYDGESLGLTALAYQADDD 505
 QY 500 EISFDPENLITGIEVXXEGGWRGYPDGHFGKXPANYVELI 539

Db 506 EISFPDDITINEMIDGMRGV-CKGRYGLFPANYVEL 544

RESULT 4
SRCB_CHICK STANDARD; PRT; 563 AA.
ID SRCB_CHICK
AC Q01406;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Src substrate protein p85 (p80) (Corractin).
GN Ems1 OR p85.25.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RX MEDLINE=92017786; PubMed=1922035;
RA Wu H., Reynolds A.B., Kammer S.B., Vines R.R., Parsons J.T.;
RT cytoskeleton-associated p85src substrate.";
RL Mol. Cell. Biol. 11:5113-5124(1991).
CC -1- FUNCTION: May contribute to the organization of cell structure.
CC The SH3 motif may function as a binding region to cytoskeleton.
CC Tyrosine phosphorylation in transformed cells may contribute to
CC cellular growth regulation and transformation.
CC -1- SUBCELLULAR LOCATION: In normal cells, probably in association
CC with the plasma membrane and possibly the endoplasmic reticulum.
CC p80/85 colocalizes with F-actin in peripheral extensions of normal
CC cells and reconstitutes (podosomes) of src-transformed cells. p80/85
CC probably associates with components of the cytoskeleton.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, p85 (shown here) and p80, may be produced by
CC alternative initiation;
CC -1- PTM: In normal cells, appears to be phosphorylated on serine and
CC threonine; in cells expressing activated forms of pp60-src, they
CC become heavily phosphorylated on tyrosine in vitro.
CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; M73705; AAA49031.1; -
DR PIR; A41530; A41530.
DR HSSP; P19174; 2HSP.
DR InterPro; IPR000134; Hs1/Corractin.
DR InterPro; IPR000108; Neu_Cyc_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF02218; Hs1_rep; 7.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KM Phosphorylation; Repeat; SH3 domain; Cytoskeleton;
KM Alternative initiation; Repeat; SH3 domain; Cytoskeleton;
FT CHAIN 1 563 SRC SUBSTRATE PROTEIN P85, ISOFORM P85.
FT INT MET 11 563 SRC SUBSTRATE PROTEIN P85, ISOFORM P80.
FT DOMAIN 92 338 6.5 X 37 AA TANDEN REPEATS.
FT REPEAT 92 128 1.
FT REPEAT 129 165 2.

FT REPEAT 166 202 3.
FT REPEAT 203 239 4.
FT REPEAT 240 276 5.
FT REPEAT 277 313 6.
FT REPEAT 314 338 7 (INCOMPLETE).
FT DOMAIN 505 563 SH3.
SQ SEQUENCE 563 AA; 63329 MW; 22A824A08B8D54C CRC64;
Query Match 5.1%; Score 204.5; DB 1; Length 563;
Best Local Similarity 36.1%; Pred. No. 1.5e-09;
Matches 52; Conservative 18; Mismatches 45; Indels 29; Gaps 5;
Db 424 PAEPAPSPX--PCLVQAEAEVYEPX-----EQETFE-----QPLVQGXKS 468
419 PAEPKPPSSRYVDANVSYSASAYKNSSTTYSAEHPESGTYTGSQYQEAVSOREAEYE 478
QY 469 EVIDHHIQGGGLSXQ-----GLCARLYDQADDDXETSPDENLITGIEYX 515
Db 479 PETTYEVAGAGDHYQAEENTYDEYENELGITALYDQAGDDELSPDDITINEMI 538
QY 516 XEGMWRGYGPDGHFGKXPANYVEL 539
Db 539 DDCWNRGV-CKGRYGLFPANYVEL 561
RESULT 5
SRCB_HUMAN STANDARD; PRT; 550 AA.
ID SRCB_HUMAN
AC 014247;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Src substrate coractin (Amplixin) (Oncogene Ems1).
GN Ems1 OR CTN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=92195667; PubMed=1532244;
RA Schuring E.M.D., Verhoeven B., Wool W.J., Michalides R.J.A.;
RT "Identification and cloning of two overexpressed genes, UZ1B3./PRAD1
RT and Ems1, within the amplified chromosome 11q13 region in human
RT carcinomas.";
RL Oncogene 7:355-361(1992).
CC (2)
CC SEQUENCE FROM N.A.
CC TISSUE=Breast;
CC MEDLINE=9233650; PubMed=8474448;
RX Schuring E.M.D., Verhoeven B., Litvinov S., Michalides R.J.A.;
RA "The product of the Ems1 gene, amplified and overexpressed in human
RA carcinomas, is homologous to a v-src substrate and is located in
RT cell-substratum contact sites.";
RL Mol. Cell. Biol. 13:2891-2898(1993).
CC -1- FUNCTION: May contribute to the organization of cell structure.
CC The SH3 motif may function as a binding region to cytoskeleton.
CC Tyrosine phosphorylation in transformed cells may contribute to
CC cellular growth regulation and transformation.
CC -1- SUBUNIT: Interacts with SHANK2 and SHANK3 via its SH2 domain (By
CC similarity).
CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----


```

RA  Meljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritze C., Holzer E., Moestl H., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Lehraich R., Reinhardt R., Pohl T.M.,
RA  Egger P., Zimmermann W., Wedler H., Rambaut R., Purnelle S.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA  Galibert F., Aves S.-J., Xiang Z., Hunt C., Moore K., Huzar S.M.,
RA  Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA  Daga R.R., Cruzado J., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA  Shavkovski G.V., Ussery D., Bartell B.G., Nurse P.;
RA  "the genome sequence of Schizosaccharomyces pombe.";
RA  Nature 415:871-880(2002).
RL  [2]
RX  SEQUENCE OF 560-857 FROM N.A.
RP  STRAIN=PR745;
RX  MEDLINE=98162722; PubMed=9501991;
RA  Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.;
RT  "Identification of open reading frames in Schizosaccharomyces pombe
RT  cDNAs.";
RL  DNA Res. 4:363-369(1997).
CC  -1- SIMILARITY: Contains 2 SH3 domains.
CC  CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-stb.ch/announce/
CC  or send an email to license@isb-stb.ch).
CC  CC
DR  EMBL; AL162631; CAB83085.1; -
DR  EMBL; D89164; BA13826.1; ALT_SEQ.
DR  HSSP; P19174; 2HSP.
DR  GeneDB SPombe; SPAP760.02c; -
DR  InterPro; IPR002108; Actbind cofin.
DR  InterPro; IPR000108; Neu_cyt_fac2_2.
DR  InterPro; IPR001452; SH3.
DR  Pfam; PF00241; co11lin_ADF. 1.
DR  Pfam; PF00018; SH3; 2.
DR  PRINTS; PR00499; P67PHOX.
DR  PRINTS; PR00452; SH3DOMAIN.
DR  ProDom; PD000066; SH3; 2.
DR  SMART; SM00102; ADF. 1.
DR  SMART; SM00326; SH3; 2.
DR  PROSITE; PS50002; SH3; 2.
KW  SH3 domain; Repeat.
KW  DOMAIN
FT  725 SH3 1.
FT  800 SH3 2.
SQ  DOMAIN
SQ  SEQUENCE 857 AA; 91137 MW; 247EFA78BAFE7097 CRC64;

Query March 3.6%; Score 143.5; DB 1; Length 857;
Best Local Similarity 27.3%; Pred. No. 0.00027;
Matches 51; Conservative 25; Mismatches 60; Indels 51; Gaps 10;

QY 387 P Q P K L R S P F L Q X O L T Q P E T F R G R P A A X X S R P A D L A - E E P A S X P P C I V Q A E E R A V Y 445
Db 689 P E A Q O L N P V A P L P P H D E T - - D E P V G - - - G D V A T H S T O P T K P A I V - - - - I Y 734

QY 446 E E P K O E F F Y Q P P L V Q - Q O X X G S E H I D - - - - H I O G - Q G L - - - - - - - - - - - - - - - - - - - - 480
Db 735 D Y S E E E N E I E - - L V E N E Q I Q I L E F V D D G M W L G E N S G Q O G L F P S N Y V E I T G P N E T A N N 791

QY 481 - - - - - S X Q G I C A R A L Y D Y O A A D D X E I S F P E N L I T G I E V X X B G M W R G Y G P D G F G K X P 533
Db 792 P P A E F Q A G P K S X V A I Y D Y O A Q E D N E L S F F E D E I I A N V D C V D P W M W E G - E C H G R G L F P 850

QY 534 A N Y V E L I 540
Db 851 S N I V E E I 857

```

ID	ABP1_YEAST	STANDARD	PRT	592 AA
AC	PI5851			
DT	01-APR-1990	(Rel. 14, Created)		
DT	01-MAY-1992	(Rel. 22, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Actin binding protein.			
GN	ABP1 OR YCR088W OR YCR88W.			
OS	Saccharomyces cerevisiae (Baker's Yeast).			
CC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetaceae; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90136906; PubMed=2405279;			
RA	Drubin D.G., Mulholland J., Zhu Z., Botstein D.;			
RT	"Homology of a yeast actin-binding protein to signal transduction			
RL	proteins and myosin-I";			
RL	Nature 343:288-290(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Frontali L., Grisanti P.;			
RL	Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: MAY BE INVOLVED IN THE SPATIAL ORGANIZATION OF CELL			
CC	SURFACE GROWTH. AN OVERPRODUCTION OF ABP1 CAUSES THE ASSEMBLY OF			
CC	THE CORTICAL ACTIN SKELETON AT INAPPROPRIATE SITES ON THE CELL			
CC	SURFACE, RESULTING IN DELocalIZED SURFACE GROWTH.			
CC	-1- SUBCELLULAR LOCATION: CORTICAL CYTOSKELETON.			
CC	-1- SIMILARITY: TO SIGNAL TRANSDUCTION PROTEINS AND MYOSIN I.			
CC	-1- SIMILARITY: Contains 1 SH3 domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.jeb-sib.ch/announce/			
CC	or send an email to license@jeb-sib.ch).			
CC	-----			
DR	EMBL; X51780; CAA36075.1; -;			
DR	EMBL; X59720; CAA42253.1; -;			
DR	PIR; S19503; LBY.			
DR	PDB; 1HQZ; 21-DEC-01.			
DR	PDB; 1J08; 01-MAR-02.			
DR	GeneOnline; 138985; -;			
DR	SGD; S0000684; ABP1.			
DR	GO; GO:0005938; C:cell cortex; IDA.			
DR	GO; GO:0005737; C:cytoplasm; IDA.			
DR	GO; GO:000515; F:protein binding; IDA.			
DR	GO; GO:0000147; P:actin cortical patch assembly; IMP.			
DR	GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IMP.			
DR	InterPro; IPR002108; Actbind,cofin.			
DR	InterPro; IPR001018; Neu_cyt_fact_2.			
DR	InterPro; IPR001452; SH3.			
DR	pfam; PF00241; cofilin_ADF; 1.			
DR	pfam; PF000318; SH3; 1.			
DR	PRINTS; PR00499; P6/PHOX.			
DR	PRINTS; PR00452; SH3DOMAIN.			
DR	ProDom; PD000066; SH3; 1.			
DR	SMART; SM00102; ADF; 1.			
DR	SMART; SM00326; SH3; 1.			
DR	PROSITE; PS50002; SH3; 1.			
KM	Cytoskeleton; Actin-binding; SH3 domain; Repeat; 3D-structure.			
FT	DOMAIN	86	97	
FT	DOMAIN	532	592	
FT	DOMAIN	200	575	
FT	REPEAT	200	209	
FT	REPEAT	436	445	
FT	REPEAT	566	575	
FT	REPEAT	58	58	
FT	CONFLICT	312	312	
FT	CONFLICT	592 AA;	63576 MW;	39525107049494AA CRC64;

Query Match 3.6%; Score 143; DB 1; Length 592;
 Best Local Similarity 27.2%; Pred. No. 0.0002;
 Matches 56; Conservative 21; Mismatches 79; Indels 50; Gaps 8;

QY SSSPPGKRSPPLOXQLOPETHFGR-----EP-----412
 DB TISKPSKQEPKAEAEQPKDYKXIONPLPGMHIEADNEEPENDDDWDEDE 445
 QY 413 AAAXSRPR-----ADLPAEED-----APSXP-----PCLVQAEAEAYEEPEQET 454
 DB 446 AAQPLPSRNVAAGAPVQKEPEQELAPSLPSRNSIPAPQOEAPQAP-EEIEEEAE 504
 QY 455 YEQPLVQOQXXGSHIDHIIQOGLSXQGLCAPALYDQADDXEISFPDENLITGEV 514
 DB 505 EAAPLPSRSSASAPPPPPRRATPEKKPKENPMWAIHVDYDAEDNEELTFVNDKINIEF 564

QY 515 XXEGWVRG-YGPDGHFGKXPANYVEL 539
 DB 565 VDDDWLGLLEKDGSKGLFPSNYVEL 590

RESULT 10
 LAST_HUMAN STANDARD; PRT; 261 AA.

AC Q14847; Q96ED2;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE LIM and SH3 domain protein 1 (LASP-1) (MLN 50).
 GN LASP1 OR MLN50.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=breast carcinoma;
 RX MEDLINE=96039245; PubMed=7490069;
 RA Lidereau R., Bassot P., Rio M.-C.;
 RA "Identification of four novel human genes amplified and overexpressed
 in breast carcinoma and localized to the q11-q21.3 region of
 chromosome 17";
 RT Genomics 28:367-376 (1995).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=23388257; PubMed=12477932;
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Suetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
 RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedon T.B., Toshiyuki S., Carninci P., Prange S.J.,
 RA Raha S.S., Lomelianno N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley A.C., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
 RA Schnerch A., Schen U.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP DOMAINS.
 RX MEDLINE=96033982; PubMed=7569475;
 RA Tomasetto C., Moog-Lutz C., Regnier C.H., Schreiber V., Bassot P.,
 RA Rio M.-C.;

RT "Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by
 the association of LIM and SH3 domains";
 RL FEBS Lett. 373:245-249 (1995).
 CC -1- SIMILARITY: Contains 1 LIM zinc-binding domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- DATABASE: NAME=Atlas Genet. Cytoenet. Oncol. Haematol.;
 WWW="http://www.infobiogen.fr/services/chronocare/Genes/lasp1ID203.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/annouce/>
 or send an email to license@isb-sib.ch).

CC EMBL; X82456; CAA57833.1; -
 DR EMBL; BC012460; AAH12460.1; -
 DR HSSP; P80171; IZFO.
 DR SWISS-2DPAGE; Q14847; HUMAN.
 DR Genew; HGNC:6513; LASP1.
 DR MIM; 602920; -
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR000900; Nebulin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00412; LIM; 1.
 DR Pfam; PF00880; Nebulin; 2.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00453; SH3DOMAIN.
 DR Prodom; PD000064; LIM; 1.
 DR Prodom; PD000066; SH3; 1.
 DR Prodom; SM00132; LIM; 1.
 DR SMART; SM00227; NEBU; 2.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00478; LIM DOMAIN 1; 1.
 DR PROSITE; PS50023; LIM DOMAIN 2; 1.
 DR PROSITE; PS50002; SH3 1.
 DR LIM domain; Metal-binding; Zinc; SH3 domain.
 FT DOMAIN 5 56
 FT DOMAIN 202 261
 FT DOMAIN 201 204
 FT CONFLICT 79 79
 FT CONFLICT 210 210
 FT CONFLICT 220 220
 SQ SEQUENCE 261 AA; 29717 MW; 3B89B988605B3639 CRC64;

Query Match 3.5%; Score 141; DB 1; Length 261;
 Best Local Similarity 29.9%; Pred. No. 0.00013;
 Matches 44; Conservative 17; Mismatches 48; Indels 38; Gaps 7;

QY 409 GREPAAXXSRP-----RADLPAEED-APSXPCLVQAEAEAV-----YEEPEQETFE 456
 DB 138 GMEBERDDSDGSSYRPLAQCPHIFPSAPVYQCPQCPVAVQSGYGVKEPA----- 190

QY 457 QPPLVQOQXXGSHIDHIIQOGLSXQGLCAPALYDQADDXEISFPDENLITGEVXX 516
 DB 191 APVSIQSRABG-----GGRKRYAVDYVSADDEVSFGDGTIVNQOIQD 236

QY 517 EGMWRCYGP---DGHGKXPANYVELI 540
 DB 237 GGMW-IGTVERTGDTGMLPANYVELI 261

RESULT 11
 ABPL_SACEX STANDARD; PRT; 617 AA.

AC P38478;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Actin binding protein.
 GN ABP1.

```

OS Saccharomyces exiguus (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCB1_TaxID=34358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 379;
RX MEDLINE=9415401; PubMed=8110838;
RA Lange U., Steiner S., Grollig F., Wagner G., Philippson P.;
RT "Cloning and sequencing of a gene coding for an actin binding protein
RT of Saccharomyces exiguus.";
RL Biochim. Biophys. Acta 1217:214-218(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN THE SPATIAL ORGANIZATION OF CELL
CC SURFACE GROWTH. AN OVERPRODUCTION OF APT1 CAUSES THE ASSEMBLY OF
CC THE CORTICAL ACTIN SKELETON AT INAPPROPRIATE SITES ON THE CELL
CC SURFACE, RESULTING IN DELOCALIZED SURFACE GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CORTICAL CYTOSKELETON.
CC -1- SIMILARITY: TO SIGNAL TRANSDUCTION PROTEINS AND MYOSIN I.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73977; CAA52156.1; -
DR FIR; S42719; S42719.
DR HSSP; P19174; ZHSP.
DR InterPro; IPR002108; Actbind_cofln.
DR InterPro; IPR001018; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; Cofilin_ADF_1.
DR Pfam; PF00018; SH3_1.
DR PRINTS; PR00459; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3_1.
DR SMART; SM00102; ADF_1.
DR SMART; SM00326; SH3_1.
DR PROSITE; PS00002; SH3_1.
DR Cytoskeleton; Actin-binding; SH3 domain; Repeat.
KM CYTOSKELETON; ACTIN-BINDING; (POTENTIAL).
FT DOMAIN 86 97
FT 557 617
FT 486 501
FT 523 529
FT 202 600
FT 202 211
FT 444 453
FT 581 600
FT 495 538
FT 510 538
FT REPEAT 495 510
FT REPEAT 523 538
FT SEQUENCE 617 AA; 68420 MW; 540A785191B65F85 CRC64;
SQ
Query Match 3.4%; Score 134.5; DB 1; Length 617;
Best Local Similarity 25.0%; Pred. No. 0.0011;
Matches 52; Conservative 21; Mismatches 82; Indels 53; Gaps 8;

```

```

OY 513 EVXHEGMRG-YGPDGHFGKXPANYEL 539
DB 588 EFVDDMWLGELETTGQKGLFPSNYVL 615
RESULT 12
ID SYN1_HUMAN STANDARD; PRT; 705 AA.
AC P17600; 075825;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Synapsin I (Brain protein 4.1).
GN SYN1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=90243651; PubMed=2110562;
RA Suedhof T.C.;
RT "The structure of the human synapsin I gene and protein.";
RL J. Biol. Chem. 265:14932-14937(1990).
RN [2]
RP SEQUENCE OF 1-125 FROM N.A.
RX MEDLINE=90368667; PubMed=2118519;
RA Sauerwald A., Hoesehe C., Oeschwald R., Kiljmann M.W.;
RT "The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and
RT CAT-less, phylogenetically conserved sequence with cell
RT type-specific promoter function.";
RL J. Biol. Chem. 265:14932-14937(1990).
RN [3]
RP SEQUENCE OF 1-258 FROM N.A.
RA Graefham D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,
CC binds to the cytoskeleton, and is believed to function in the
CC regulation of neurotransmitter release.
CC -1- SUBCELLULAR LOCATION: Synapse.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Synapsin IA;
CC IsoId=PI7600-1; Sequence=Displayed;
CC Name=Synapsin IB;
CC IsoId=PI7600-2; Sequence=VSP_006316, VSP_006317;
CC -1- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC SYNAPSIN I IN THE NERVE TERMINAL.
CC -1- SIMILARITY: Belongs to the synapsin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58378; AAC41930.1; -
DR EMBL; M58321; AAC41930.1; JOINED.
DR EMBL; M58341; AAC41930.1; JOINED.
DR EMBL; M58351; AAC41930.1; JOINED.
DR EMBL; M58353; AAC41930.1; JOINED.
DR EMBL; M58359; AAC41930.1; JOINED.
DR EMBL; M58371; AAC41930.1; JOINED.
DR EMBL; M58372; AAC41930.1; JOINED.
DR EMBL; M58373; AAC41930.1; JOINED.
DR EMBL; M58374; AAC41930.1; JOINED.
DR EMBL; M58375; AAC41930.1; JOINED.
DR EMBL; M58375; AAC41930.1; JOINED.
DR EMBL; M58377; AAC41930.1; JOINED.

```

463 QOXXSEH 470
550 AAAGCPH 657

Db

RESULT 13

ID MYT1_FHUMAN STANDARD; PRT; 1098 AA.

AC 000160; Q8WMN7; 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Myosin I ϵ (Myosin-IE).

GN MYO1F

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxID=9606;

RN RN

RN SEQUENCE FROM N.A.

RX MEDLINE=21664126; PubMed=11804589;

RA Krugmann S., Anderson K.E., Ridley S.H., Rizzo N., McGregor A., Coedwell J., Davidson K., Equitao A., Ellison C.D., Lipp P., Manfellat M., Katschke N., Painter G., Thuring J.W., Cooper M.A., Lim Z.-Y., Holmes A.B., Dove S.K., Mitchell R.H., Grewal A., Nazarian A., Erdjument-Bromage H., Tempst P., Stephens L.R., Hawkins P.T.;

RA "Identification of ARAP3, a novel PI3K effector regulating both Arp and Rho Gases, by selective capture on phospholipid affinity matrices.";

RT Mol. Cell 9:95-108(2002).

RL

RN

RN SEQUENCE FROM N.A.

RC TISSUE=Spleen;

RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y., Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y., Sugiyama T., Irie R., Otsubki T., Saito H., Ota T., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wakatsushima M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;

RA "NDSO human cDNA sequencing project.";

RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RL

RN

RN [3]

RN SEQUENCE FROM N.A.

RC TISSUE=Blood;

RX MEDLINE=22386257; PubMed=12477932;

RA Klausberg R.L., Faingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bueltow K.H., Scheffer C.F., Hsieh N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Bash F., Ditchenko L., Mansina K., Farmer A.A., Rubin G.M., Hong L., Stjepanovic M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udell T.B., Tohyuki S., Abramson R.D., Mullaly S.J., Rana S.S., Lquallano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bonak S.A., McKeown P.J., McKernan K.U., Malek J.A., Guaratne P.H., Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Faley J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield J.S.N., Krzywnicki M.I., Skalska U., Smalls D.E., Schnerch A., Schein U.B., Jones S.J.W., Maira V.A., "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RL

RN

RN [4]

RN SEQUENCE OF 318-1038 FROM N.A.

RC TISSUE=Retina;

RX MEDLINE=97237053; PubMed=9119401;

RA Crozet F., El Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.

RA Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
 RA Weil D., Pujol R., Petit C.;
 RT "Cloning of the genes encoding two murine and human cochlear
 RT unconventional type I myosins";
 RL Genomics 40:332-341(1997).
 CC -I- FUNCTION: Myosins are actin-based motor molecules with ATPase
 CC activity. Unconventional myosins serve in intracellular movements.
 CC Their highly divergent tails are presumed to bind to membranous
 CC compartments, which would be moved relative to actin filaments (By
 CC similarity).
 CC -I- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
 CC -I- SIMILARITY: Contains 1 IQ domain.
 CC -I- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ310570; CAC83948.1; -;
 DR EMBL: AK092877; BAC03995.1; -;
 DR EMBL: BC028071; AAH28071.1; -;
 DR EMBL: X98411; CAA67058.1; -;
 DR HSSP: P08799; 1MND.
 DR Genew: HGNC:17600; MYO1F.
 DR MIM: 601480; -;
 DR GO: GO:0016461; C:unconventional myosin; NAS.
 DR GO: GO:0003779; F:actin binding; NAS.
 DR GO: GO:0005524; F:ATP binding; NAS.
 DR GO: GO:0005516; F:calmodulin binding; NAS.
 DR InterPro: IPR000048; IQ region.
 DR InterPro: IPR001609; myosin_head.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PRO0193; MYOSINHEAVY.
 DR PRINTS: PRO0452; SH3DOMAIN.
 DR ProDom: PD000355; myosin_head; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00242; MYSC; 1.
 DR SMART: SM00325; SH3; 1.
 DR PROSITE: PS50096; IQ; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
 KW Multigene family.
 KM -----
 FT DOMAIN 1 677
 FT 693 722
 FT 1041 1088
 FT 579 589
 FT 110 117
 FT 259 259
 FT 534 536
 FT 592 593
 FT 602 602
 FT 759 759
 FT 797 797
 FT 805 805
 FT 809 809
 FT 814 814
 FT 922 922
 FT 927 927
 FT 930 930
 FT 948 948
 FT 958 958
 FT 971 971
 FT 1031 1031
 FT 1098 1098
 FT 1098 AA; 124803 MW; B7FCID371C421B8 CRC64;
 SQ SEQUENCE

Query Match 3.2%; Score 129.5; DB 1; Length 1098;
 Best Local Similarity 21.1%; Pred. No. 0.0052; Indels 103; Gaps 10;
 Matches 58; Conservative 23; Mismatches 91;
 QY 285 GLXINAKSKESRPQXITRQRRGASVGLGXEXKRLSPAPGPPHAPRRGRARAPRAGT 344
 DB 908 GRLLTVSVGGLPKRSKPTKRG--MAKGRRESQAFTAPAA-PPRGMDRNVPPSARG 964
 QY 345 GVRGXPPFVLGNARPERRRXKPPSPRTASWGWRTYXSTSSPQPKLSPPFQXLTLP 404
 DB 965 G---PLPLRTMSGGTTRPPRPPS-----TSIQASRRPPARP-----SH 1003
 QY 405 ETRF-----GREPAAXSRPADLPAPBPAPSPXPCLVQAEBAVBE 447
 DB 1004 NTEFLNVPDQMGAMGRKRSVGRFPVGVGRPK---PQRTHGRRC----- 1046
 QY 448 PXQETTYEOPPLVQOQXXSHIDHIGQGLSKXGLCARALYDQADDXISPPEN 507
 DB 1047 -----RALYQYVGQDVDELSPNVNE 1066
 QY 508 LITGIEYXEGWNRG--YGPGRFGKXPANYVELI 540
 DB 1067 VITLMEDEPSGWMKGRLLHGEGLF---PGNYVEKI 1098
 RESULT 14
 NEBU HUMAN STANDARD; PRT; 6669 AA.
 ID NEBU HUMAN
 AC P20928; O15346;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nebulin.
 GN NEB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN RN
 RX SEQUENCE FROM N.A.
 RX MEDLINE=95257391; PubMed=7739042;
 RA Label S., Kolmerer B.;
 RT "The complete primary structure of human nebulin and its correlation
 RT to muscle structure";
 RL J. Mol. Biol. 248:308-315(1995).
 RP PARTIAL PRELIMINARY SEQUENCE:
 RX MEDLINE=88284704; PubMed=3397062;
 RA Zeviani M., Darras B.T., Rizzuto R., Salvati G., Betto R.,
 RA Bonilla E., Miranda A.F., Du J., Samit C., Dickson G., Walsh F.S.,
 RA Dimauro S., Francke U., Schon E.A.;
 RT "Cloning and expression of human nebulin cDNAs and assignment of the
 RT gene to chromosome 2q31-q32";
 RL Genomics 2:249-256(1988).
 [3]
 RP STRUCTURE BY NMR OF 6610-6669.
 RX MEDLINE=98179559; PubMed=9514727;
 RA Poltoun A.S., Millevoi S., Gaudel M., Kolmerer B., Pastore A.;
 RT "SH3 in muscles: solution structure of the SH3 domain from nebulin";
 RL J. Mol. Biol. 276:189-202(1998).
 CC -I- FUNCTION: This giant muscle protein may be involved in maintaining
 CC the structural integrity of sarcomeres and the membrane system
 CC associated with the myofibrils. Bind and stabilize F-actin.
 CC -I- TISSUE SPECIFICITY: Muscle specific. Located in the thin filament
 CC of striated muscle.
 CC -I- DISEASE: Defects in NEB are a cause of the autosomal recessive
 CC form of nemaline myopathy (NEM2).
 CC -I- SIMILARITY: Contains 1 SH3 domain.
 CC -I- SIMILARITY: Contains 178 nebulin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-stb.ch/announce/> or send an email to license@isb-stb.ch).

EMBL; X83957; CAA58788.1; ALT_SEQ.
EMBL; M19668; AAA59916.1; ALT_SEQ.
EMBL; M19669; AAA59917.1; ALT_SEQ.
PIR; A29979; A29979.
PIR; A29979; A29979.
PIR; S55024; S55024.
PDB; INEB; 28-VAN-98.
GeneW; HGNC:7720; NEB.
MIM; 256030; -.
GO; GO:0015629; C:actin cytoskeleton; TAS.
GO; GO:0030017; C:barcomere; NAS.
GO; GO:0008307; P:structural constituent of muscle; TAS.
GO; GO:0008307; P:structural constituent of actin filament length; NAS.
GO; GO:0007525; P:somatic muscle development; NAS.
InterPro; IPR000900; Nebulin.
InterPro; IPR001452; SH3.
Pfam; PF00880; Nebulin; 146.
Pfam; PF00018; SH3; 1.
PRINTS; PR00510; NEBULIN.
PRINTS; PR00452; SH3DOMAIN.
ProDom; PD000066; SH3; 1.
PROSITE; PS50002; SH3; 1.
Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.

REPEAT 76 107 NEBULIN 1.
REPEAT 113 143 NEBULIN 2.
REPEAT 148 178 NEBULIN 3.
REPEAT 183 213 NEBULIN 4.
REPEAT 218 248 NEBULIN 5.
REPEAT 253 283 NEBULIN 6.
REPEAT 289 318 NEBULIN 7.
REPEAT 324 354 NEBULIN 8.
REPEAT 363 393 NEBULIN 9.
REPEAT 398 428 NEBULIN 10.
REPEAT 434 464 NEBULIN 11.
REPEAT 502 532 NEBULIN 12.
REPEAT 537 567 NEBULIN 13.
REPEAT 573 603 NEBULIN 14.
REPEAT 611 641 NEBULIN 15.
REPEAT 681 711 NEBULIN 16.
REPEAT 749 779 NEBULIN 17.
REPEAT 784 814 NEBULIN 18.
REPEAT 820 850 NEBULIN 19.
REPEAT 858 888 NEBULIN 20.
REPEAT 893 923 NEBULIN 21.
REPEAT 924 954 NEBULIN 22.
REPEAT 959 990 NEBULIN 23.
REPEAT 993 1023 NEBULIN 24.
REPEAT 1028 1058 NEBULIN 25.
REPEAT 1064 1094 NEBULIN 26.
REPEAT 1102 1132 NEBULIN 27.
REPEAT 1137 1167 NEBULIN 28.
REPEAT 1168 1198 NEBULIN 29.
REPEAT 1204 1234 NEBULIN 30.
REPEAT 1237 1267 NEBULIN 31.
REPEAT 1272 1302 NEBULIN 32.
REPEAT 1308 1338 NEBULIN 33.
REPEAT 1346 1376 NEBULIN 34.
REPEAT 1381 1411 NEBULIN 35.
REPEAT 1412 1442 NEBULIN 36.
REPEAT 1448 1478 NEBULIN 37.
REPEAT 1481 1511 NEBULIN 38.
REPEAT 1516 1546 NEBULIN 39.
REPEAT 1552 1582 NEBULIN 40.
REPEAT 1590 1620 NEBULIN 41.
REPEAT 1625 1655 NEBULIN 42.

REPEAT 1656 1686 NEBULIN 43.
REPEAT 1692 1722 NEBULIN 44.
REPEAT 1725 1755 NEBULIN 45.
REPEAT 1760 1790 NEBULIN 46.
REPEAT 1796 1826 NEBULIN 47.
REPEAT 1834 1864 NEBULIN 48.
REPEAT 1869 1899 NEBULIN 49.
REPEAT 1900 1930 NEBULIN 50.
REPEAT 1936 1966 NEBULIN 51.
REPEAT 1969 1999 NEBULIN 52.
REPEAT 2004 2034 NEBULIN 53.
REPEAT 2040 2070 NEBULIN 54.
REPEAT 2078 2108 NEBULIN 55.
REPEAT 2113 2143 NEBULIN 56.
REPEAT 2144 2174 NEBULIN 57.
REPEAT 2180 2210 NEBULIN 58.
REPEAT 2213 2243 NEBULIN 59.
REPEAT 2248 2278 NEBULIN 60.
REPEAT 2284 2314 NEBULIN 61.
REPEAT 2322 2352 NEBULIN 62.
REPEAT 2357 2387 NEBULIN 63.
REPEAT 2388 2418 NEBULIN 64.
REPEAT 2423 2453 NEBULIN 65.
REPEAT 2456 2486 NEBULIN 66.
REPEAT 2491 2521 NEBULIN 67.
REPEAT 2527 2557 NEBULIN 68.
REPEAT 2565 2595 NEBULIN 69.
REPEAT 2600 2630 NEBULIN 70.
REPEAT 2631 2661 NEBULIN 71.
REPEAT 2666 2696 NEBULIN 72.
REPEAT 2699 2729 NEBULIN 73.
REPEAT 2734 2764 NEBULIN 74.
REPEAT 2770 2800 NEBULIN 75.
REPEAT 2808 2838 NEBULIN 76.
REPEAT 2843 2873 NEBULIN 77.
REPEAT 2874 2904 NEBULIN 78.
REPEAT 2909 2939 NEBULIN 79.
REPEAT 2942 2972 NEBULIN 80.
REPEAT 2977 3007 NEBULIN 81.
REPEAT 3013 3043 NEBULIN 82.
REPEAT 3051 3081 NEBULIN 83.
REPEAT 3086 3116 NEBULIN 84.
REPEAT 3117 3147 NEBULIN 85.
REPEAT 3152 3182 NEBULIN 86.
REPEAT 3185 3215 NEBULIN 87.
REPEAT 3220 3250 NEBULIN 88.
REPEAT 3256 3286 NEBULIN 89.
REPEAT 3294 3324 NEBULIN 90.
REPEAT 3329 3359 NEBULIN 91.
REPEAT 3360 3390 NEBULIN 92.
REPEAT 3395 3425 NEBULIN 93.
REPEAT 3428 3458 NEBULIN 94.
REPEAT 3463 3493 NEBULIN 95.
REPEAT 3499 3529 NEBULIN 96.
REPEAT 3537 3567 NEBULIN 97.
REPEAT 3572 3602 NEBULIN 98.
REPEAT 3603 3633 NEBULIN 99.
REPEAT 3638 3668 NEBULIN 100.
REPEAT 3671 3701 NEBULIN 101.
REPEAT 3706 3736 NEBULIN 102.
REPEAT 3742 3772 NEBULIN 103.
REPEAT 3780 3810 NEBULIN 104.
REPEAT 3815 3845 NEBULIN 105.
REPEAT 3846 3876 NEBULIN 106.
REPEAT 3894 3924 NEBULIN 107.
REPEAT 3949 3979 NEBULIN 108.
REPEAT 3984 4014 NEBULIN 109.
REPEAT 4021 4052 NEBULIN 110.
REPEAT 4057 4087 NEBULIN 111.
REPEAT 4088 4118 NEBULIN 112.
REPEAT 4123 4153 NEBULIN 113.
REPEAT 4156 4186 NEBULIN 114.
REPEAT 4191 4220 NEBULIN 115.

```

FT REPEAT 4226 4256 NEBULIN 116.
FT REPEAT 4264 4294 NEBULIN 117.
FT REPEAT 4299 4329 NEBULIN 118.
FT REPEAT 4330 4360 NEBULIN 119.
FT REPEAT 4365 4395 NEBULIN 120.
FT REPEAT 4400 4430 NEBULIN 121.
FT REPEAT 4435 4465 NEBULIN 122.
FT REPEAT 4471 4501 NEBULIN 123.
FT REPEAT 4544 4574 NEBULIN 124.

Query Match 3.2%; Score 126; DB 1; Length 6669;
Best Local Similarity 27.0%; Pred. No. 0.069;
Matches 47; Conservative 13; Mismatches 72; Indels 42; Gaps 5;

QY 402 TQPTHTFRRPAAAXSRP-----ADLPAEPAPSPPCVLQAEF-----442
DB 6503 TNPQSVFDPYDAEDNIQGRSLHMTNVQAKRRSRQSSASALSVSGGEKSEHSEADPH 6562
QY 443 -----AYEEFEEQETFEYEPQVQOXXGSEHIDHIIQGGSLXQGLCARA 489
DB 6563 LSTVSDGCVFAVSTAYKAKTTELPQCRSSSVATQCTVSSIPSH-----PSTAGKIFRA 6617
QY 490 LYDQADDXEISPDENLITGIEVXXEGMWRGYP---DGHFGKXPANYELI 540
DB 6618 MYDYMAADADBVSPFGDPAIINVQAIQEGMW--YGTVQRTGRTMLPANYEAL 6669

RESULT 15
ID ITN1 RAT STANDARD; PRT; 1217 AA.
AC Q9WVE9; Q9WVE1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Intersectin 1 (EH domain and SH3 domain regulator of endocytosis 1).
GN ITSN1 OR ITSN OR EHSN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=99303609; PubMed=10373452;
RA Okamoto M., Schoch S., Suedhof T.C.;
RT "EHSN1/Intersectin, a protein that contains EH and SH3 domains and
RT binds to dynamin and SNAP-25. A protein connection between exocytosis
RT and endocytosis?";
RU J. Biol. Chem. 274:18446-18454 (1999).
CC -1- FUNCTION: Adapter protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters
CC several dynamin in a manner that is regulated by alternative
CC splicing. Also binds clathrin-associated proteins and other
CC components of the endocytic machinery, such as N-WASP, Epsins and
CC Stecin 2 (By similarity)".
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC Enriched in synaptosomes.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9WVE9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9WVE9-2; Sequence=VSP_004297;
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
CC -1- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,
CC bind to dynamin.
CC -1- DOMAIN: The KIERO domain binds to SNAP-25 and SNAP-23.
CC -1- SIMILARITY: Contains 2 EH-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 5 SH3 domains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF127798; AAD30271.1; -
CC DR EMBL; AF132672; AAD31026.1; -
CC DR HSSP; P29354; IGFC.
CC DR InterPro; IPR002048; EF-hand.
CC DR InterPro; IPR000261; Eps15_homology.
CC DR InterPro; IPR001452; SH3.
CC DR Pfam; PF00036; ehnd; 3.
CC DR Pfam; PF00018; SH3; 5.
CC DR PRINTS; PR00452; SH3DOMAIN.
CC DR ProDom; PD000066; SH3; 5.
CC DR SMART; SM00054; Eph; 2.
CC DR SMART; SM00027; EH; 2.
CC DR SMART; SM00326; SH3; 5.
CC DR PROSITE; PS00018; EF_HAND; 2.
CC DR PROSITE; PS00031; EH; 2.
CC DR PROSITE; PS00002; SH3; 5.
CC KM Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
CC Alternative splicing.
CC FT DOMAIN 21 109
CC FT CA_BIND 66 78 EH 1.
CC FT DOMAIN 221 310 EF_HAND 1 (POTENTIAL).
CC FT CA_BIND 267 279 EF_HAND 2 (POTENTIAL).
CC FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KIERO).
CC FT DOMAIN 350 670 COILED COIL (POTENTIAL).
CC FT DOMAIN 737 803 SH3 1.
CC FT DOMAIN 910 968 SH3 2.
CC FT DOMAIN 968 999 SH3 3.
CC FT DOMAIN 1057 1057 SH3 4.
CC FT DOMAIN 1071 1135 SH3 5.
CC FT DOMAIN 1152 1211 POLY-SER.
CC FT DOMAIN 321 324 Missing (in isoform 2).
CC FT VARSPLIC 1003 1073 /FTId=VSP_004297.
CC SQ SEQUENCE 1217 AA; 137154 MW; 6C13238A5A5B34B CRC64;

Query Match 3.1%; Score 124.5; DB 1; Length 1217;
Best Local Similarity 23.0%; Pred. No. 0.015;
Matches 41; Conservative 19; Mismatches 47; Indels 71; Gaps 5;

QY 368 PSRT-----ARSGTWRTXYSISSPOGKLRSPPLOXULTQPTHTFGRPPAXXSRPRAD 422
DB 855 PSTNKEPTEDWDITAAQPSLTVPSAGLR-----QRSFTPATATGSSRSP-----902
QY 423 LPAEPAPSPPCVLQAESEAVYEBPXEQETFEYEPQVQOXXGSEHIDHIIQGGSLX 482
DB 903 -----VLQGGKRV 910
QY 483 QGLCARALYDQADDXEISPDENLITGIEVXXEGMWRGYPDGHFGKXPANYELI 540
DB 911 EQLQALYPRPAKNDHNLNFKKSDVITVLB-QQDMWFG-EVQQKGMFRRSYKLI 966

```

Search completed: April 27, 2004, 10:53:37
 Job time : 16.5828 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:45:34 ; Search time 42.2397 Seconds
(without alignments)
5923.479 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989
Sequence: 1 HEIPVPRVYPAKPOXKRA.....DFKSKKKKKLKLMTSSX 793

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	796	20.0	302	4	Q9HBN8
2	796	20.0	430	4	Q9JUB6
3	796	20.0	430	4	Q9NR72
4	796	20.0	431	4	Q9EF30
5	794	19.9	439	4	Q9EK74
6	628.5	15.8	433	11	Q62418
7	628.5	15.8	436	11	Q60WPI
8	627.5	15.7	432	11	Q8BH56
9	606	15.2	429	11	Q9JW74
10	603.5	15.1	432	11	Q9JW66
11	602	15.1	433	11	Q9JW67
12	599.5	15.0	436	11	Q9JH4
13	365.5	9.2	447	13	Q7ZXC9
14	231	5.8	559	5	Q9YDF4
15	228.5	5.7	486	11	Q9Z218
16	227.5	5.7	384	5	Q8MRS9

17	227.5	5.7	531	5	Q9VUB4	Q9VUB4 drosophila
18	226.5	5.7	643	5	Q9XUT0	Q9XUT0 caenorhabd
19	223	5.6	559	5	Q96046	Q96046 drosophila
20	210	5.3	530	13	Q8UWC3	Q8UWC3 xenopus lae
21	209	5.2	471	11	Q70419	Q70419 rattus norv
22	209	5.2	508	11	Q70420	Q70420 rattus norv
23	208	5.2	509	11	Q92116	Q92116 mus musculu
24	208	5.2	509	11	Q8BNAS	Q8BNAS mus musculu
25	207.5	5.2	477	5	Q96E20	Q96E20 suberites d
26	205	5.1	513	4	Q96H99	Q96H99 homo sapien
27	201	5.0	587	5	Q96459	Q96459 strongyloce
28	184.5	4.6	206	13	Q7SXH4	Q7SXH4 brachydanio
29	168	4.2	758	5	Q6EK16	Q6EK16 dictyostell
30	167.5	4.2	634	4	Q8N707	Q8N707 homo sapien
31	160.5	4.0	513	4	Q9NWB9	Q9NWB9 homo sapien
32	153.5	3.8	239	4	Q13802	Q13802 homo sapien
33	152	3.8	475	4	Q13801	Q13801 homo sapien
34	148.5	3.7	390	4	Q13249	Q13249 homo sapien
35	148	3.7	401	4	Q13147	Q13147 homo sapien
36	147	3.7	543	11	Q35823	Q35823 rattus norv
37	146	3.7	263	11	Q99M28	Q99M28 rattus norv
38	146	3.7	368	3	Q9P837	Q9P837 candida alb
39	146	3.7	475	4	Q9BV70	Q9BV70 homo sapien
40	145	3.6	263	6	Q77506	Q77506 oryctolagus
41	143	3.6	315	13	Q03292	Q03292 xenopus lae
42	140.5	3.5	473	5	Q9Y0S9	Q9Y0S9 drosophila
43	136.5	3.4	447	5	Q86K04	Q86K04 dictyostell
44	134.5	3.4	1098	11	Q8CG29	Q8CG29 mus musculu
45	134.5	3.4	1098	11	Q811E7	Q811E7 mus musculu

ALIGNMENTS

RESULT 1

Q9HBN8 PRELIMINARY; PRT; 302 AA.

AC Q9HBN8; ID Q9HBN8; DT 01-MAR-2001 (TEMBUREL. 16, Created)
DT 01-MAR-2001 (TEMBUREL. 16, Last sequence update)
DT 01-MAR-2003 (TEMBUREL. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J.R., Man D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF18020; AAG17262.1; --
DR HSBP; P23727; 1PNT
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KM Hypothetical protein; SH3 domain.
SQ SEQUENCE 302 AA; 3438 MW; 2DEGALC7511612 CRC64;

Query Match 20.0%; Score 796; DB 4; Length 302;
Best Local Similarity 82.5%; Pred. No. 1.1e-70;
Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY RPERRPXRRSRTARSGTWKXTXSSPPGKLRSPFLQXOTOPETRGRRPAXXSR 418
DB 120 RNEQSAVHREIFPKQERAMSTTSISPPGKLRSPFLQXOTOPETRGRRPAXXSR 179

QY 419 PRADIPAEPPAPSPXPCLVQAEAEVYEPKQETFEYQPPVVOOQXXGSEHIDHIQO 478
 DB 180 PRADIPAEPPAPSPXPCLVQAEAEVYEPKQETFEYQPPVVOOQXXGSEHIDHIQO 239
 QY 479 GLSXQGLCARALYDYQADDXEISPDENLITGIEVXXEGWWRGYPDGHFGKXPANYVE 538
 DB 240 GLSXQGLCARALYDYQADDXEISPDENLITGIEVDEGWRGYPDGHFGKXPANYVE 299
 QY 539 LIE 541
 DB 300 LIE 302
 RESULT 2
 Q9UT06 PRELIMINARY; PRT; 430 AA.
 ID Q9UT06
 AC Q9UT06
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE SRC homolog 3 domain-containing protein HIP-55 (Drebrin F).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX MEDLINE=20036526; PubMed=10567356;
 RA Eisenat D., Yao Z., Wang X.-S., Kori R., Zhou G., Lee S.C., Tan T.-H.;
 RT "A Novel Src Homology 3 Domain-containing Adaptor Protein, HIP-55,
 RT that Interacts with Hematopoietic Progenitor Kinase 1.";
 RL J. Biol. Chem. 274:33945-33950(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Zhang M., Yuan Z., Wan T., He L., Gao X.;
 RT "Molecular cloning of cDNA encoding drebrin F";
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Straubeberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF197060; AAF13701.1; -
 DR EMBL; AF077353; AAF80228.1; -
 DR EMBL; BC031687; AAF31687.1; -
 DR HSBP; P23727; IPNU.
 DR GO; GO:0008047; F:enzyme activator activity; TAS.
 DR GO; GO:0005515; P:protein binding; TAS.
 DR GO; GO:0007257; P:activation of JUN; TAS.
 DR InterPro; IPR002108; Actbind_cofin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; Cofilin_ADP; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR ProSITE; PSS0002; SH3; 1.
 SQ SEQUENCE 430 AA; 48207 MW; 7B8C42ED047257AE CRC64;
 Query Match 20.0%; Score 796; DB 4; Length 430;
 Best Local Similarity 82.5%; Pred. No. 1.6e-70;
 Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;
 QY 359 RPERPXXRPSRTARSWGTXYSISSPGKLRSPFLQOLQPEPTHFGRPAAXSR 418
 DB 248 RNEGSADVHPRERIKOKERAMSTTSSISPOGKRSFLOKOLQPEPTHFGRPAAXSR 307
 QY 419 PRADIPAEPPAPSPXPCLVQAEAEVYEPKQETFEYQPPVVOOQXXGSEHIDHIQO 478
 DB 308 PRADIPAEPPAPSPXPCLVQAEAEVYEPKQETFEYQPPVVOOQXXGSEHIDHIQO 367
 QY 479 GLSXQGLCARALYDYQADDXEISPDENLITGIEVXXEGWWRGYPDGHFGKXPANYVE 538

DB 368 GLSXQGLCARALYDYQADDXEISPDENLITGIEVDEGWRGYPDGHFGKXPANYVE 427
 QY 539 LIE 541
 DB 428 LIE 430
 RESULT 3
 Q9NR72 PRELIMINARY; PRT; 430 AA.
 ID Q9NR72
 AC Q9NR72
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cervical SH3p7 (Mucin-associated protein).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX MEDLINE=99077625; PubMed=9858486;
 RA Gipsom I.K., Spurr-Michaud S., Moccia R., Zhan O., Toribara N.,
 RA Ho S.B., Garduno A.R., Hill J.A. III;
 RT "MUC4 and MUC5B transcripts are the prevalent mucin messenger
 RT ribonucleic acids of the human endocervix";
 RL Biol. Reprod. 60:58-64(1999).
 CC -1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; AF250287; AAF81273.1; -
 DR EMBL; AF151364; AAG3120.1; -
 DR HSBP; P23727; IPNU.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003779; P:actin binding; IEA.
 DR InterPro; IPR002108; Actbind_cofin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; Cofilin_ADP; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00102; ADP; 1.
 DR SMART; SM00326; SH3; 1.
 DR ProSITE; PSS0002; SH3; 1.
 KW SH3 domain.
 SQ SEQUENCE 430 AA; 48154 MW; 7D0CAF17D8CEB8DC CRC64;
 Query Match 20.0%; Score 796; DB 4; Length 430;
 Best Local Similarity 82.5%; Pred. No. 1.6e-70;
 Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;
 QY 359 RPERPXXRPSRTARSWGTXYSISSPGKLRSPFLQOLQPEPTHFGRPAAXSR 418
 DB 248 RNEGSADVHPRERIKOKERAMSTTSSISPOGKRSFLOKOLQPEPTHFGRPAAXSR 307
 QY 419 PRADIPAEPPAPSPXPCLVQAEAEVYEPKQETFEYQPPVVOOQXXGSEHIDHIQO 478
 DB 308 PRADIPAEPPAPSPXPCLVQAEAEVYEPKQETFEYQPPVVOOQXXGSEHIDHIQO 367
 QY 479 GLSXQGLCARALYDYQADDXEISPDENLITGIEVXXEGWWRGYPDGHFGKXPANYVE 538
 DB 368 GLSXQGLCARALYDYQADDXEISPDENLITGIEVDEGWRGYPDGHFGKXPANYVE 427
 QY 539 LIE 541
 DB 428 LIE 430

RESULT 4
ID 096F30 PRELIMINARY; PRT; 431 AA.
AC 096F30;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to src homology 3 domain-containing protein HRP-55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; BC011677; AH01677.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; F:actin_bind.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofilin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 431 AA; 48294 MW; D18P9D316FFD0B4E CRC64;

Query Match 20.0%; Score 796; DB 4; Length 431;
Best Local Similarity 82.5%; Pred. No. 1.6e-70;
Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 REPRPPXKPRSTARSWGTWRTXKSSPQPKLSPFOKXOLTOPETHFGREPAAXSR 418
DB 248 RNEQSAVHPREIFKQERAMSTSISSPQKLSPLQKLTPEPHFGREPAALSR 308
QY 419 PRADLPAREPAPSPPCIVQAEAEAVYEEPEQETFFYQPLVQOQXXGSEHIDHIIQO 478
DB 309 PRADLPAREPAPSPPCIVQAEAEAVYEEPEQETFFYQPLVQOQXXGSEHIDHIIQO 368
QY 479 GISXQGLCARALYDQADDXEISPDPELITGIEVXXEGMWRGYPDGHFGMPANYVE 538
DB 369 GISXQGLCARALYDQADDXEISPDPELITGIEVXXEGMWRGYPDGHFGMPANYVE 428
QY 539 LIE 541
DB 429 LIE 431

RESULT 5
ID 096K74 PRELIMINARY; PRT; 439 AA.
AC 096K74;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein FL114461.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Magatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi K., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Matanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niimura K., Iwayanagi T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK027367; BA853065.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; F:actin_bind.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofilin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3; 1.
KW Hypothetical protein; SH3 domain.
SQ SEQUENCE 439 AA; 49042 MW; 8C044FBD0B2C6D5 CRC64;

Query Match 19.3%; Score 794; DB 4; Length 439;
Best Local Similarity 91.3%; Pred. No. 2.6e-70;
Matches 148; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 381 TSISSPQPKLSPFOKXOLTOPETHFGREPAAXSRPRADLPAREPAPSPPCIVQAE 440
DB 279 TSISSPQPKLSPFOKXOLTOPETHFGREPAALSRPRADLPAREPAPSPPCIVQAE 338
QY 441 EBAVYEEPEQETFFYQPLVQOQXXGSEHIDHIIQOGLSXQGLCARALYDQADDXE 500
DB 339 EBAVYEEPEQETFFYQPLVQOQXXGSEHIDHIIQOGLSXQGLCARALYDQADDXE 398
QY 501 ISFPDENLITGIEVXXEGMWRGYPDGHFGMPANYVEIE 541
DB 399 ISFPDENLITGIEVXXEGMWRGYPDGHFGMPANYVEIE 439

RESULT 6
ID 062418 PRELIMINARY; PRT; 433 AA.
AC 062418;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Drebilin-like SH3 domain-containing protein SH3P7.
GN DBIL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=98234438; PubMed=9630982;
RL Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.,
CC -1- SIMILARITY: CONTAINS A COPY OF THE SH3 DOMAIN.
CC -1- SIMILARITY: TO DREBRIN.
DR EMBL; U58884; AAC52640.1; -
DR HSP; P19174; 2HSP.
DR MGI; MGI:700006; Dbn1.
DR GO; GO:0005737; C:cytosol; IEA.
DR GO; GO:0030227; C:lamellipodium; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0016601; P:RAC protein signal transduction; IEA.
DR InterPro; IPR002108; Actbind_cofln.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofilin_ADF; 1.
DR Pfam; PF00018; SH3; 1.

DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000066; SH3; 1.
 DR SMART: SM00102; ADF; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00002; SH3; 1.
 KW SH3 domain.
 FT DOMAIN 356 432 SH3.
 SQ SEQUENCE 433 AA; 48428 MW; 602D3862C446PAA0 CRC64;

Query Match 15.8%; Score 628.5; DB 11; Length 433;
 Best Local Similarity 58.7%; Pred. No. 7.7e-54;
 Matches 131; Conservative 12; Mismatches 43; Indels 37; Gaps 5;

QY 336 AAPPPAGTGVRCXVPVLGNGARPERPXXRPSRTASWGT-----W 378
 DB 231 AGAPSRGT-----EPEQAV--SRTQWESAGQAAPHREIFKOKERA 272
 QY 379 RXTXSISSPQKLSRPFLOXQLOPETHFGREPAAXSRPADLPAREPAPSPCTVQ 438
 DB 273 KSTTSVTSQPKLSPFLQKQLOPETHFGREPAAPSPACV-CEEPSTLS-AQ 330
 QY 439 AEEBAVEEPXQETFFYQPPVQQQXXGSHDHIIQGGLSXQGLCAPALYDQAAD 498
 DB 331 TEEEPYVEPPEODLYEPPVQQQAGSEHIDIMYQSGQGLCAPALYDQAAD 390
 QY 499 XEISPPENTLTGIEVXXEGMWRGYPDGHGKXPANVVELI 541
 DB 391 TEISFDPENLTGIVIDEGMWRGYPDGHGKXPANVVELI 433

RESULT 7
 Q80WP1 PRELIMINARY; PRT; 436 AA.

AC 080WP1; 080WP1; 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Actin-binding protein 1.
 GN ABP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL6/J;
 RC Hou P., Estrada L., Gorecki J.D.,
 RT "Fgdl, a Cdc42-specific guanine nucleotide exchange factor, directly
 RT interacts with cortactin and Ablp1 to modulate cell shape".
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY098595; AA028340.1;
 DR GO; GO:0005623; C:intracellular; IEA.
 DR GO; GO:0003779; F:actin binding; IEA.
 DR InterPro; IPR002108; Actbind_cofan.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; Cofilin_ADF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00102; ADF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00002; SH3; 1.
 SQ SEQUENCE 436 AA; 48699 MW; 85AEP9781C698A3F CRC64;

Query Match 15.8%; Score 628.5; DB 11; Length 436;
 Best Local Similarity 55.2%; Pred. No. 7.8e-54;
 Matches 133; Conservative 15; Mismatches 58; Indels 35; Gaps 5;
 QY 318 LSPPRGPPPLAPRRGAGAPAPAGTGVRCXVPVLGNGARPERPXXRPSRTASWGT 377
 DB 214 LQEAARRQRYQDQHRSGAPSPSRGT-----EPQQAV--SRTQWES 257
 QY 378 -----WRXTXSISSPQKLSRPFLOXQLOPETHFGREPAAXSRPR 420

DB 258 AGQAAPHREIFKOKERAMSTTSVTSQPKLSPFLQKQLOPETHFGREPAAPSPRA 317
 QY 421 ADLPAREPAPSPCTVQAEBAVEEPXQETFFYQPPVQQQXXGSHDHIIQGGL 480
 DB 318 AGV-CEEPAPSTLS-AQTEEPYVEPPEODLYEPPVQQQAGSEHIDIMYQSGQ 375
 QY 481 SXQGLCAPALYDQAADDEXISPPENTLTGIEVXXEGMWRGYPDGHGKXPANVVELI 540
 DB 376 SGQGLCAPALYDQAADDTISFDPENLTGIVIDEGMWRGYPDGHGKXPANVVELI 435
 QY 541 E 541
 DB 436 E 436

RESULT 8
 Q8BH56 PRELIMINARY; PRT; 432 AA.
 AC 08BH56; 08BH56; 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Drexlin-like.
 GN BDNF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain, Eye, and Medulla oblongata;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA "The RIKEN Genome Exploration Research Group Phase I & II Team."
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs".
 RL Nature 420:563-573(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,
 RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Millaby S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Straubeberg R.;
 RL Submitted (Feb-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK046073; BAC32592.1; -
 DR EMBL; AK053796; BAC35528.1; -
 DR EMBL; AK078082; BAC37118.1; -
 DR EMBL; BC046430; AAH46430.1; -
 DR MGD; MGI:700006; Ddn1.
 DR GO; GO:0005737; C:Cytoplasm; IDA.

DR GO:0030027; C:lamellipodium; IDA.
 DR GO:0003779; F:actin binding; IDA.
 DR GO:0005515; F:protein binding; IDA.
 DR GO:0016601; P:Rac protein signal transduction; IDA.
 DR InterPro; IPR002108; Actbind_cofilin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; cofilin_ADF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00102; ADF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 SQ SEQUENCE 432 AA; 48341 MW; 4A36D3B6E59C707F CRC64;

Query Match 15.7%; Score 627.5; DB 11; Length 432;
 Best Local Similarity 58.0%; Pred. No. 9,7e-54;
 Matches 130; Conservative 13; Mismatches 44; Indels 37; Gaps 5;

QY 335 RARAPRAGTGVRCXVPVLGNGARPERRPXKRSRTARSGT----- 377
 DB 229 RSAGAPRTG-----EPQEAIV---SRTQEWESAGQAHPHREIFKQKER 270
 QY 378 WXTXSSISGPOPKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLV 437
 DB 271 AMSTSVSVSSQPKLRSPFLQKLTQPTHTFGRETAVSRRPAGV-CEELAPSTPS-A 328
 QY 438 QAEEAVYEEPEXQETFFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALYDQAD 497
 DB 329 QRESEPTVEPPEOQTLVEEPPLVQOXXGSEHIDHINYSQFSGQGLCARALYDQAD 388
 QY 498 DXEISPDENLITGIEVXXEGMWRGYPDGHFGMKPANYVELIE 541
 DB 389 TGEIVDEGMWRGYPDGHFGMKPANYVELIE 432

RESULT 9

Q9JMT4 PRELIMINARY; PRT; 429 AA.

AC Q9JMT4; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE SH3P74 (Fragment).
 GN SH3P74.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamazaki H., Shitao T.;
 RT "Rattus norvegicus mRNA for SH3P7 (SH3P74).";
 RC Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; AB038364; BAA90866.1; -.
 DR HSSP; P19174; ZHSP.
 DR GO:0005622; C:intracellular; IEA.
 DR GO:0003779; F:actin binding; IEA.
 DR InterPro; IPR002108; Actbind_cofilin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; cofilin_ADF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00102; ADF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 KW SH3 domain.
 SQ SEQUENCE 429 AA; 47923 MW; A7E7E7BBAF4A1705 CRC64;
 FT NON TER 1
 Query Match 15.2%; Score 606; DB 11; Length 429;

Best Local Similarity 59.9%; Pred. No. 1.3e-51;
 Matches 127; Conservative 13; Mismatches 58; Indels 14; Gaps 5;

QY 335 RARAPRAG-----TGVRCXVPVLGNGARPERRPXKRSRTARSGTWRXTXSSISGPOP 389
 DB 227 RSAGAPRTGELGEVSVRSRQWESAGQA-PPHREIFKQKERAMS-----TTSVSSSQP 279
 QY 390 GKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLVQAEEAVYEEFX 449
 DB 280 GKLRSPFLQKLTQPTHTFGRETAVSRRPAGV-CEELAPSTPS-AQTDEPTVEPS 337
 QY 450 EDETFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALYDQADDXISFDPENLI 509
 DB 338 EDETFYEQPPLVQOXXGSEHIDHINYSQFSGQGLCARALYDQADDTGISFDPENLI 397
 QY 510 TGEIVXXEGMWRGYPDGHFGMKPANYVELIE 541
 DB 398 TGEIVDEGMWRGYPDGHFGMKPANYVELIE 429

RESULT 10

Q9JMT6 PRELIMINARY; PRT; 432 AA.

AC Q9JMT6; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE SH3P73.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamazaki H., Shitao T.;
 RT "Rattus norvegicus mRNA for SH3P7 (SH3P73).";
 RC Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; AB039619; BAA92709.1; -.
 DR HSSP; P19174; ZHSP.
 DR GO:0005622; C:intracellular; IEA.
 DR GO:0003779; F:actin binding; IEA.
 DR InterPro; IPR002108; Actbind_cofilin.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00241; cofilin_ADF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00102; ADF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 KW SH3 domain.
 SQ SEQUENCE 432 AA; 48254 MW; E2D5E5C1D325B641 CRC64;

Query Match 15.1%; Score 603.5; DB 11; Length 432;
 Best Local Similarity 59.0%; Pred. No. 2.4e-51;
 Matches 125; Conservative 13; Mismatches 61; Indels 13; Gaps 4;

QY 335 RARAPRAG-----TGVRCXVPVLGNGARPERRPXKRSRTARSGTWRXTXSSISGPOP 389
 DB 229 RSAGAPRTGELGEVSVRSRQWESAGQA-PPHREIFKQKERAMS-----TTSVSSSQP 282
 QY 390 GKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLVQAEEAVYEEFX 449
 DB 283 GKLRSPFLQKLTQPTHTFGRETAVSRRPAGV-CEELAPSTPS-AQTDEPTVEPS 340
 QY 450 EDETFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALYDQADDXISFDPENLI 509
 DB 341 EDETFYEQPPLVQOXXGSEHIDHINYSQFSGQGLCARALYDQADDTGISFDPENLI 400
 QY 510 TGEIVXXEGMWRGYPDGHFGMKPANYVELIE 541
 DB 401 TGEIVDEGMWRGYPDGHFGMKPANYVELIE 432

```

RESULT 11
O9JUN67 PRELIMINARY; PRT; 433 AA.
AC O9JUN67;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE SH3P7r2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB039818; BA92708.1; -.
DR HSSP; P19174; ZHSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofillin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 433 AA; 48341 MW; 952PFF81E6A6EDDA CRC64;

Query Match 15.1%; Score 602; DB 11; Length 433;
Best Local Similarity 58.9%; Pred. No. 3,4e-51;
Matches 123; Conservative 12; Mismatches 44; Indels 30; Gaps 4;

QY 361 ERPRXKPSRT-----ARSWGTR-----XTXSISSPQGL 392
DB 227 QHSAGPSPKRGELGEVVSRSROEMESAGQAPHPREIFKQERAMSTTSVSSQPKL 286
QY 393 RSPFLQXQLTOPETHFGREPAAXXSRPADLPABEPAPSPCLVQAEBAVYEEPEXQ 452
DB 287 RSPFLQXQLTOPETHFGREPAAXXSRPADLPABEPAPSPCLVQAEBAVYEEPEXQ 344
QY 453 TFYQPPPLVQOQXGSEHIDHIOGGLSXQGLCARALYDYQADDEKISFDPENLITGI 512
DB 345 TLYEPPVQVQPGAGSGHIDYVQSGGLCARALYDYQADDEKISFDPENLITGI 404
QY 513 EVXKEGWRGYPGDPGHFGKXPANYVELIE 541
DB 405 EVIDEGWWRGYPGDPGHFGKXPANYVELIE 433

RESULT 12
O9JUN67 PRELIMINARY; PRT; 436 AA.
AC O9JUN67;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE SH3P7r1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB039818; BA92708.1; -.
DR HSSP; P19174; ZHSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofillin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 436 AA; 48612 MW; 064CA1B6A84444 CRC64;

Query Match 15.0%; Score 599.5; DB 11; Length 436;
Best Local Similarity 56.4%; Pred. No. 6e-51;
Matches 127; Conservative 12; Mismatches 47; Indels 39; Gaps 5;

QY 354 LGNGARPERR-----PXKRPSPRT-----ARSWGTR----- 379
DB 214 LGNAARQRQYQEGHRSAGPSPSPKRGELGEVVSRSROEMESAGQAPHPREIFKQ 273
QY 380 ---XTXSISSPQGLKSPFLQXQLTOPETHFGREPAAXXSRPADLPABEPAPSPCL 436
DB 274 RAMSTTSVSSSQPKLSPFLQXQLTOPETHFGREPAAXXSRPADLPABEPAPSPCL 331
QY 437 VQAEBAVYEEPEXQETFEYEQPLVQOQXGSEHIDHIOGGLSXQGLCARALYDYQAA 496
DB 332 AQTDEPTEVPSSEOTLIEPFPVQPGAGSGHIDYVQSGGLCARALYDYQAA 391
QY 497 DDXEISFDPENLITGIEVXSGWRGYPGDPGHFGKXPANYVELIE 541
DB 392 DDEISFDPENLITGIEVIDEGWWRGYPGDPGHFGKXPANYVELIE 436

RESULT 13
O7ZXQ9 PRELIMINARY; PRT; 447 AA.
AC O7ZXQ9;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to drebrin-like.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo.
RA Klein S., Strausberg R.;
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC044296; AAH44296.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofillin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.

```

```

RT "Rattus norvegicus mRNA for SH3P7 (SH3P7r1)."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT "SH3P7."
RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB038365; BA90867.1; -.
DR EMBL; AB039346; BA90819.1; -.
DR HSSP; P19174; ZHSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofillin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 436 AA; 48612 MW; 064CA1B6A84444 CRC64;

Query Match 15.0%; Score 599.5; DB 11; Length 436;
Best Local Similarity 56.4%; Pred. No. 6e-51;
Matches 127; Conservative 12; Mismatches 47; Indels 39; Gaps 5;

QY 354 LGNGARPERR-----PXKRPSPRT-----ARSWGTR----- 379
DB 214 LGNAARQRQYQEGHRSAGPSPSPKRGELGEVVSRSROEMESAGQAPHPREIFKQ 273
QY 380 ---XTXSISSPQGLKSPFLQXQLTOPETHFGREPAAXXSRPADLPABEPAPSPCL 436
DB 274 RAMSTTSVSSSQPKLSPFLQXQLTOPETHFGREPAAXXSRPADLPABEPAPSPCL 331
QY 437 VQAEBAVYEEPEXQETFEYEQPLVQOQXGSEHIDHIOGGLSXQGLCARALYDYQAA 496
DB 332 AQTDEPTEVPSSEOTLIEPFPVQPGAGSGHIDYVQSGGLCARALYDYQAA 391
QY 497 DDXEISFDPENLITGIEVXSGWRGYPGDPGHFGKXPANYVELIE 541
DB 392 DDEISFDPENLITGIEVIDEGWWRGYPGDPGHFGKXPANYVELIE 436

RESULT 13
O7ZXQ9 PRELIMINARY; PRT; 447 AA.
AC O7ZXQ9;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to drebrin-like.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo.
RA Klein S., Strausberg R.;
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC044296; AAH44296.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; cofillin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.

```

DR SMART; SM00102; ADF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 SQ SEQUENCE 447 AA; 50611 MW; 38C6767EE6E3B521 CRC64;

Query Match 9.2%; Score 365.5; DB 13; Length 447;
 Best Local Similarity 45.8%; Pred. No. 1.2e-27;
 Matches 81; Conservative 19; Mismatches 54; Indels 23; Gaps 6;

QY 386 SPQSGKLRSPFLQXQLTPET-----HFGREPAAXSRERADLPABEAP--SXP-- 434
 DB 273 SAQGRRLRSPFLQXACQSPSPSPRHRYQDEPPSPVYAHQTPESPVPVSHPPS 332
 QY 435 -----CLV-QAEBAVYEEPEXDEFTFYEOPPLVQ--QXGSEHIDHIGQGLSXG 484
 DB 333 TVHWKQCTASQOEENITQDATEDQNTYEDTTNNTYEDTPQEEPYEIEVE--EKG 390
 QY 485 LCARALYQYQADADXEISFDPENITIGIEVXXEGMWRGYPGDGHGKMPANYVELLE 541
 DB 391 VCARALYQYQADDEISFPDDLITQIFDEGMWGFSPAGHGMFPANYVELLE 447

RESULT 14
 Q9VDF4 PRELIMINARY; PRT; 559 AA.

AC Q9VDF4 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Contractin protein (LD29964P).
 GN CONTRACTIN OR CG3637.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
 Amaratunga P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Barton R.C., Rogers J.H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch L., Bessey E.M.,
 Bailew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
 Beeson K.Y., Bencze P.V., Bernier J.P., Bhattacharya D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 Burris K.J., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durkin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.D., Hernandez J.R., Houck J.,
 Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalish J., Karpen G.H., Ke Z., Kennison J.A., Keithum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Palazotto M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 Wang Z.-Y., Wastaman D.A., Weinstein G.M., Weissbach U.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Abmayyan A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Fartan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacle J., Paragas V., Park S., Pounanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; A0003733; AAF55840.1;
 DR EMBL; AY051774; AAK93198.1;
 DR HSSP; P29355; ISEM.
 DR FlyBase; FBgn0025865; Contractin.
 DR InterPro; IPR003134; Hs1/Contractin.
 DR InterPro; IPR00108; Neu_cyt_fact_2.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF02218; Hs1_rep; 4.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00499; P67PHOX.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 KW SH3 domain.
 SQ SEQUENCE 559 AA; 61083 MW; 2B767BD3EFF1BCB2 CRC64;

Query Match 5.8%; Score 231; DB 5; Length 559;
 Best Local Similarity 27.1%; Pred. No. 4e-14;
 Matches 77; Conservative 43; Mismatches 110; Indels 54; Gaps 12;

QY 291 SKESRP--GTRQGRGASVGLGXEXRL-----SPQGRPLHAPKRG----- 334
 DB 292 TSTAPPKSGRAITGRGGIGNALISAPNMQSVSTPPAKKPIITIPKQPKIEL 351
 QY 335 RARAPPAGTGVACXVPVLGNAR--PERRPXXRPSRTASWGTWXTYSSISPPQKIR 393
 DB 352 EAKEPPTASTSAVAAPTPVVPARAPETAPVAKAAPDPVQIETVTPRPSRQ 411
 QY 394 SPFLQXQLTPETPHFGREPAAXSRERADLPABEAPSPKPCUYQAEBAVYEEPEXQET 453
 DB 412 SP-VVYPTPQPEVH-----AQVQVQ-----EPQPADPEPV-VVEEPLYNQAEIKR 457
 QY 454 FYEQP-----LVQOQX--XGSEHIDHIGQGLSXGGLCARALYQYQAA 496
 DB 458 ASPLPFPNGVSAVAAPSGTATVPEEAYVANSNDLVLEDTGTH-----AIALYQYQAA 512
 QY 497 DXEISFDPENITIGIEVXXEGMWRGYPGDGHGKMPANYVELI 540
 DB 513 DDEISFPDDVITHIEKIDGWMRGKCN-RYGLFPANYQV 555

RESULT 15
 Q92218 PRELIMINARY; PRT; 486 AA.

AC Q92218 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hematopoietic cell specific lym substrate 1.
 GN HCSL1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Strauberg R.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases
 DR EMBL, BC007469, AA007469.1, -.
 DR MED, MG1:104568, Hc18.
 DR GO, GO:0017124, F3:SH3-domain binding, IDA.
 DR InterPro, IPRO03143, Hs1/Contactin.
 DR InterPro, IPRO0108, Neu_cyt_fact_2.
 DR InterPro, IPRO01452, SH3
 DR Pfam, PF02218, Hs1_rep; 4.
 DR Pfam, PF00018, SH3_1.
 DR PRINTS, PR00499, P67P0X.
 DR PRINTS, PR00452, SH3DOMAIN.
 DR ProDom, PD000066, SH3_1.
 DR SMART, SM00326, SH3; 1.
 DR PROSITE, PSS0002; SH3; 1.
 SQ SEQUENCE 466 AA; 54240 MW; 0EE14FEFA0A31412 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:23:34 ; Search time 53.1947 Seconds

(without alignments)
4164.275 Million cell updates/sec

Title: US-10-028-952a-10

Perfect score: 3752 1 IRHEHGETEVXNKXKAPG.....DXALALQPKXERTLSQKX 784

Sequence: 1 IRHEHGETEVXNKXKAPG.....DXALALQPKXERTLSQKX 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Genesep29Jan04:*
2: genesep1980s:*
3: genesep1990s:*
4: genesep2000s:*
5: genesep2001s:*
6: genesep2002s:*
7: genesep2003as:*
8: genesep2003bs:*
9: genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3251.5	86.7	753	4	AA666395 Human mam
2	1177.5	31.4	760	4	AA666394 Human pro
3	435	11.6	538	2	AA23756 Human mam
4	238	6.3	55	3	AA653794 Human col
5	199	5.3	40	3	AA653737 Human col
6	194.5	5.2	91	3	AA653743 Human col
7	185	4.9	73	3	AA653817 Human col
8	149	4.0	7052	6	AA635266 Human P45
9	145	3.9	54	4	AA653774 Human col
10	145	3.9	54	4	AA653774 Human col
11	138	3.7	50	3	AA653773 Human col
12	122	3.3	123	7	ADC95956 Human rep
13	122	3.3	123	7	ADC95956 Human rep
14	120	3.2	96	4	AA694351 Human rep
15	120	3.2	96	4	AA694351 Human rep
16	120	3.2	96	4	AA694351 Human rep
17	120	3.2	96	4	AA694351 Human rep
18	119	3.2	97	5	ABP38827 Staphyloc
19	119	3.2	97	5	ABP38827 Staphyloc
20	119	3.2	97	5	ABP38827 Staphyloc
21	119	3.2	97	5	ABP38827 Staphyloc
22	119	3.2	97	5	ABP38827 Staphyloc
23	117	3.1	106	5	ABP38827 Staphyloc
24	116	3.1	139	3	AA653822 Human col
25	115	3.1	47	3	AA653894 Human col

26	113.5	3.0	19938	6	AB698398 Streptomy
27	113	3.0	82	4	AA685047 Human imm
28	113	3.0	86	2	AA685047 Human imm
29	111	3.0	673	6	AA684009 Human POM
30	110.5	2.9	80	4	AA684574 Human imm
31	110	2.9	465	2	AA682653 Human cal
32	110	2.9	2087	4	AA631516 Human aci
33	109.5	2.9	19938	6	ABP76681 Streptomy
34	109.5	2.9	19938	6	ABP76681 Streptomy
35	108	2.9	2091	3	AA612000 Human pro
36	106.5	2.8	476	4	ABG03460 Novel hum
37	106.5	2.8	476	4	ABG03460 Novel hum
38	106	2.8	51	6	ABG03864 Human mus
39	106	2.8	51	6	ABG03864 Human mus
40	105.5	2.8	70	4	AA692686 Human dig
41	105	2.8	221	4	AA693924 Human pro
42	105	2.8	19938	6	ABP76679 Streptomy
43	104.5	2.8	1212	2	AA687503 Human N-m
44	104	2.8	42	4	AA610628 Human pol
45	103.5	2.8	51	4	AA686866 Human imm

ALIGNMENTS

RESULT 1
ID AAB66395 standard; protein; 753 AA.

AC AAB66395;
XX 10-APR-2001 (first entry)
DT XX Human mammatatin C protein sequence.

DE Human mammatatin C protein sequence.
XX Human; mammatatin A; mammatatin B; mammatatin C; ECGI; cancer;
KW epithelial cell growth inhibitor.

OS Homo sapiens.

PN WO200076955-A1.

PD 28-DEC-2000.

XX 19-JUN-2000; 2000MO-US016900.

XX 18-JUN-1999; 99US-0139995P.

XX (BIOT-) BIOTHERAPIES INC.

XX Ervin PR;

XX WPI; 2001-071393/08.

PT A Mammatatin-like epithelial cell growth inhibitor, and the nucleic that encodes it, useful for diagnosing and/or preventing epithelial cell cancers, e.g. of the ovaries or prostate.

PS Disclosure; Page 44-46; 55pp; English.

XX The present invention describes a mammatatin-like epithelial cell growth inhibitor (ECGI) which has substantial similarity to mammatatin A, B or C. ECGI is expressed in healthy cells, but is either absent or expressed at reduced levels in cancerous cells. The protein and its coding sequence can be used to inhibit epithelial cell growth and the amount present in cells can be used to diagnose cancer or monitor its progress

XX Sequence 753 AA;

Query Match 86.7%; Score 3251.5; DB 4; Length 753;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 31; Gaps 31;

```

QY 1 IRHEGEETVEYNNKKEAFGAPVPSPRGARGRRPCPPYKXHYSDXHXDTPVRGGEP 60
DB 1 IRHEGEETVEYNNK- EAPGAPVPSPRGARGRRPCGPVYKHYSR- XTDPVRGGEP 58
QY 61 RGLAAGARPARRRPGATRSXGSAWGYLGRYTQCTQXQVSKCELRDRLLPMKRRPAK 120
DB 59 RGLA- GAKRPARRRPGATRSXGSAWGYLGRYTQCT- XQVSKCELRDRLLPM- RRAK 115
QY 121 ARLLIFSTNTDESASRSFDPFGYAXXVRRKTTGTGLMRPSXHSVAAXSDVSS 180
DB 116 ARLLIFSTNTDESASRSFDPFGYAX- XVRKTTGTGLMRPS- HSDVAXSDVSS 173
QY 181 YHAEFTKRWIHPILIGXSSWYTVVROVSFTLLMCCCHGNPAQYENRERLAVYL 240
DB 174 YHAEFTKRWIHPILIG- XSWYTVVROVSFTLL- MCCCHGNPAQYENRERLAVYL 231
QY 241 GXGANGAKXXSVGLXINASKSESFPYGTIRORRGASVGLGXDPRLSPAGRPPTSTXX 300
DB 232 GXGANGAK- XSVGLXINASKSESFP- GTIRORRGASVGLGXDPRLSPAGRPPTSTX- X 288
QY 301 RAGGRVPRAPRGSGXAECCSSWETGRGRKGYPILARHAPYRARAEFYSSGTIHRHTS 360
DB 289 RAGGRVPRAPRGSGS- AECSSWETGRGRKGYPILARHAPYRARAEF- XSTTIHRHTS 346
QY 361 ACIFVXXXLILFLMVDIQXMDXXTWXFFWFIKESYXXXSLYKFTSLPSPDFKXERYV 420
DB 347 ACIFM- XLIILFLMVDIQXMDXXTWXFFWFIKESY- XXXSLYKFTSLPSPDFK- ERYV 403
QY 421 WRAPHXHYPPFTYLLQNDPFGYRYLOVSKXROIEYVNCIRGDFYIOSCMKDKCSRDL 480
DB 404 WRAPHXHYPPFTYLLQNDPFGYRYLOV- XROIEYVNCIRGDF- IOSCMKDKCSRDL 461
QY 481 QSNMWSOMKXISSSTSSXSTELALXSLIPYXXXXXGFI SNILXGIKIXXVYLFSL 540
DB 462 QSNMWSOMKXISSSTSS- XSTELALXSLIPY- XXXXGFI SNILXGIKIXXVYLFSL 519
QY 541 AEPFONIKXXPSIGHYCTRHCVCHSKFPMWCSQXFCRVKXSLVTVLPSGRNLTXX 600
DB 520 AEPFONIK- XPSIGHYCTRHCVCH- SKFPMWCSQXFCRVKXSLVTVLPSGRNLT- X 576
QY 601 FNRKXSNRRTYXXITLQISPPHTASTCACLIGSCYFPFYLXXTTFFSPHPS 660
DB 577 FNRKXSNRRTYXIT- LQISPPHTASTCACLIGSCYFPFYLST- XTTPFSPHPS 634
QY 661 FFIIVYITNTCLSEQLIXHRRXXSTGEXXLIPIVILALXAKAGRLBSRVNDQEPQHE 720
DB 635 FFIIV- YITNTCLSEQLIXHRRXXSTGEXXLIPIVILAL- XAKAGRLBSRVNDQEP- QHE 691
QY 721 TLSIQKRTKIXPYVLAHTCLSYSEGXXIXDAOEVEAAVAXDAIALQFEXERETLS 780
DB 692 TLSIQKRTKIXPYVLAHTCLSYSEGXX- XIDAEVEAAVAXD- AIALQFEXERETLS 749
QY 781 QKXK 784
DB 750 QKXK 753

```

RESULT 2
AAB6394
ID AAB6394 standard; protein; 760 AA.

AC AAB6394;
XX
DT 10-APR-2001 (first entry)
XX Human prostate ECGI protein sequence.
XX Human; mammaplastin A; mammaplastin B; mammaplastin C; ECGI; cancer;
XX epithelial cell growth inhibitor.
XX Homo sapiens.
XX
PN WO200078955-A1.

```

XX 28-DEC-2000.
FD
XX 19-JUN-2000; 2000MO-US016900.
XX
PF 19-JUN-2000; 2000MO-US016900.
XX
PR 18-JUN-1999; 99US-0139995P.
XX
PA (BIOT-) BIOTHERAPIES INC.
XX
PI Eryn PR;
XX
DR WPI; 2001-071393/08.
XX
DR N-PSDB; AAF31282.
XX
PT A Mammaplastin-like epithelial cell growth inhibitor, and the nucleic that
PT encodes it, useful for diagnosing and/or preventing epithelial cell
PT cancers, e.g. of the ovaries or prostate.
PS Disclosure; Page 41-43; 55pp; English.
XX
CC The present invention describes a mammaplastin-like epithelial cell growth
CC inhibitor (ECGI) which has substantial similarity to mammaplastin A, B or
CC C. ECGI is expressed in healthy cells, but is either absent or expressed
CC at reduced levels in cancerous cells. The protein and its coding sequence
CC can be used to inhibit epithelial cell growth and the amount present in
CC cells can be used to diagnose cancer or monitor its progress
XX
SQ Sequence 760 AA;

```

Query Match 31.4%; Score 1177.5; DB 4; Length 760;
Best local similarity 81.7%; Pred. No. 1.9e-120;
Matches 267; Conservative 4; Mismatches 39; Indels 17; Gaps 14;

```

QY 5 HGEETVEYNNKKEAFGAPVPSPRGARGRRPCPPYKXHYSDXHXDTPVRGGEPGAL 64
DB 37 HGEETVEYNNKKEAFGAPVPSPRGARG- XRPCPPYKHYSR- FTDVRRGGEPGAL 94
QY 65 AXGAKRPARRRPGATRSXGSAWGYLGRYTQCTQXQVSKCELRDRLLPMKRRPAK 124
DB 95 ASXAKRPARRRPGATRSXGSAWGYLGRYTQCT- QVSKCELRDRLLPMKRRPAK 152
QY 125 LIFSTNTDESASRSFDPFGYAXXVRRKTTGTGLMRPSXHSVAAXSDVSSYHXE 164
DB 153 LIFSTNTD- XSGASRSFDPFGYAX- GVRKTTGTGLMRPSXSDVAFSFDVSSYHX- X 209
QY 185 AEFTKRWIHPILIGXSSWYTVVROVSFTLLMCCCHGNPAQYENRERLAVYL 244
DB 210 AEFTKRWIHPILIG- NSMDTVVROVSFTLLMCCCHGNPAQYENRER- XHLVYVVLGXGA 267
QY 245 NGAXKXSVGLXINASKSESFPYGTIRORRGASVGLGXDPRLSPAGRPPTSTXX 303
DB 268 NGAK- LSVGLXINASKSESFPYGTIRORRGASVGLGX- RLSPAGRPPLHARRGA- 324
QY 304 GRVPRARPGSGXAECCSSWETGRGRK 330
DB 325 RAEPRA--GRVRCXVYFVLGANGAR 347

```

RESULT 3
AA23756
ID AA23756 standard; protein; 538 AA.

AC AA23756;
XX
DT 09-SEP-1999 (first entry)
XX Human mammaplastin amino acid sequence.
XX Human; mammaplastin; breast cancer.
XX Homo sapiens.
XX
PN WO9332625-A2.

XX 01-JUL-1999.
 PD 18-DEC-1998; 98MO-US027147.
 XX 19-DEC-1997; 97US-00994076.
 XX (UNMI) UNITV MICHIGAN.
 PA
 PI
 DR WPI: 1999-418924/35.
 DR N-PSDB; AAB55833.
 XX
 PT DNA encoding Mammatstatin, a specific mammary cell growth inhibitor,
 PT useful for therapy of breast cancer.
 PS
 XX Claim 2; Page 45-47; 77pp; English.
 CC The present sequence represents human mammatstatin. Mammatstatin produced
 CC by normal human mammary cells in culture and recombinant expressed
 CC Mammatstatin inhibit the growth of human mammary carcinoma cells, and is
 CC useful as a therapeutic agent in the treatment of breast cancer
 SQ
 Sequence 538 AA;

Query Match 11.6%; Score 435; DB 2; Length 538;
 Best Local Similarity 28.8%; Pred. No. 1.1e-38;
 Matches 182; Conservative 50; Mismatches 216; Indels 184; Gaps 30;

QY 123 LILPSTINTDSESGASRSPFGFAXXVRYKVTGTTGLMPSPXHXSDVAFKSPDVGSSYH 182
 DB 1 MILIFSTINTDSESGASRSPFGFAXXVRYKVTGTTGLMPSPXHXSDVAFKSPDVGSSYH 182
 QY 183 XEAEFTKXWIV-----HPLIGXXSVXVTV-----RQVS----- 211
 DB 34 RD-----NWLVAAKRSKRRLILRCRLFLSLGRIHQALDCSPTNRERELGLDRREGL 88
 QY 212 FTLLMCCCHGNPAQYERNRXRLVYVLGXGANGAKXXSVGLKLNKSKSRSPXGTIRQ 271
 DB 89 FTLLMCCCHGNPAQYERNRXRLVYVLGXGANGAKLPVGLLNKSKSRSPXGTIRQ 148
 QY 272 RRGASVGLGXPRPRLSPAGRPBPSTRXKXAGRVPRAPGPGSABCPSSWETGRGK 331
 DB 149 RRGASVGLGXPRPRLSPAGRPBPSTRXKXAGRVPRAPGPGSABCPSSWETGRGK 331
 QY 332 GXPLARH-----APHVRAAEFXXSSTIHNHTSACIFMXXXILFLVNDIQWDCX 382
 DB 182 ACPAARRRRGVRSALRPCKGAKAAGKAAPSPVTHRTFVLVPSARVAPFTIDIQV-HVS 240
 QY 383 XTWKFYFWF-----IEKSSYXXXXRLYKFTSLSPDFCKXERWVWNAAPHXP-PPTXL 434
 DB 241 LLVNEFFSFGFISSGGIARSPGSSISGLLR--NLHTDFH-----RGCTVLAHYQVIFL 291
 QY 435 LQNFQKGRVYQVEXXKQIEYXNFCIRGT---DXYIOSCMKDKCSDDLQSYMKSGMY 491
 DB 292 NMKESGSEMP-LSLPLPLPYCRMTSRGTGYKPHYTDKNTEISALEAQILGK----- 345
 QY 492 ISSSTTSXXST-----ELALXSLI--FTY-----XXXXXGFIISNLI 528
 DB 346 VVGRTRISALGCTCKAGIGNLRLANTFLVVPAYIILNMLCDHLLYLITNEKGLVQIYFKV 405
 QY 529 IKIKXXVVLFSIAPFQNIKXXPSIGHY-----CTHCCHXSKXKSPWCSQ 576
 DB 406 KKSNNRSCFLHFFKTKNSYRLG--LYVDPVTFTFOCKS--HGVVHNSV-GW--- 458
 QY 577 XFCRVKSLTVRLFSKSNLTXKXENLRKYVSNRTRTYXLTITQISPYHTAS--TCACXL 634
 DB 459 -----RNSLTIVRLF--RNETSEPM-----SSTCAKLVGTGLPEPESHRSHTTLAAH 506
 QY 635 IPGSGYF-PFYFLSLXKTTSPSPHPSFPLIV 665
 DB 507 VVILFLAPKLFPPSSSHNPPFPL-L-FLEFI 537

RESULT 4

AB53794
 ID AAB53794 standard; protein; 55 AA.

AC AAB53794;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1334.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardiostatic; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.

OS Homo sapiens.

PN WC0200055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US005883.

PR 12-MAR-1999; 99US-0124270P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587534/55.

DR N-PSDB; AAC98551.

PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer.

PS Claim 11; Page 1915; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular,
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins may
 CC also be used to prevent diseases such as neural disorders, immune system
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal
 CC disorders, wounds, renal disorders, infectious diseases, and
 CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
 CC sequences used in the exemplification of the present invention
 XX

SO Sequence 55 AA;

Query Match 6.3%; Score 238; DB 3; Length 55;
 Best Local Similarity 86.5%; Pred. No. 3.5e-18;
 Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 37 CGPPVYKHYSDXXTDPVVRGEPFGALAXGAKRPAARPGATRSQASRWG 88
 DB 2 CGPPVYKHYSDXXTDPVVRGEPFGALAXGAKRPAARPGATRSQASRWG 53

RESULT 5

AB53737
 ID AAB53737 standard; protein; 40 AA.

```

XX AC AAB53737;
XX DT 09-MAR-2001 (first entry)
XX DE Human colon cancer antigen protein sequence SEQ ID NO:1277.
XX DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX DE identification; cytostatic; cardiostatic; neuroprotective; vulnary;
XX DE immunomodulatory; muscular; gynaecological; gastrointestinal;
XX DE nephrotoxic; antineoplastic; antibacterial; gene therapy; wound;
XX DE neutral disorder; immune system disorder; muscular disorder;
XX DE reproductive disorder; gastrointestinal disorder; renal disorder;
XX DE infectious disease; cardiovascular disorder.
XX OS Homo sapiens.
XX PN WO200055351-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005883.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI: 2000-587534/55.
XX DR N-PSDB; AAC98494.
XX PT Colon cancer associated gene sequences, referred to as colon cancer
XX PT antigens, useful for the treatment, prevention, and diagnosis of colon
XX PT disorders such as colon cancer.
XX PS Claim 11; Page 1864; 2104pp; English.
XX AC AAG97991 to AAC98763 encode the human colon cancer associated proteins,
XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX CC human colon cancer antigens can have cytostatic, cardiostatic, muscular,
XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX CC vulnary, nephrotoxic, antineoplastic, antibacterial activities, and
XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,
XX CC proteins and antibodies to the proteins are useful for the prevention,
XX CC treatment and diagnosis of colon disorders, such as colon cancer. The
XX CC polynucleotides may be used in diagnostics and research, such as for
XX CC chromosome identification, and as hybridisation probes. The proteins may
XX CC also be used to prevent diseases such as neural disorders, immune system
XX CC disorders, muscular disorders, reproductive disorders, gastrointestinal
XX CC disorders, wounds, renal disorders, infectious diseases, and
XX CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
XX CC sequences used in the exemplification of the present invention
XX SQ
XX Sequence 40 AA;
XX Query Match 5.3%; Score 199; DB 3; Length 40;
XX Best Local Similarity 90.0%; Pred. No. 4.8e-14;
XX Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 202 WXTVWQVSTLTLMCCCHGNPAQYERNRRLVTVG 241
DB 1 WVTIVROVSFTLTMCCCHGNPAQYERNRRLVTVG 40

```

```

DE DE Human colon cancer antigen protein sequence SEQ ID NO:1283.
XX DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX DE identification; cytostatic; cardiostatic; neuroprotective; vulnary;
XX DE immunomodulatory; muscular; gynaecological; gastrointestinal;
XX DE nephrotoxic; antineoplastic; antibacterial; gene therapy; wound;
XX DE neutral disorder; immune system disorder; muscular disorder;
XX DE reproductive disorder; gastrointestinal disorder; renal disorder;
XX DE infectious disease; cardiovascular disorder.
XX OS Homo sapiens.
XX PN WO200055351-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005883.
XX PR 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI: 2000-587534/55.
XX DR N-PSDB; AAC98500.
XX PT Colon cancer associated gene sequences, referred to as colon cancer
XX PT antigens, useful for the treatment, prevention, and diagnosis of colon
XX PT disorders such as colon cancer.
XX PS Claim 11; Page 1868-1869; 2104pp; English.
XX AC AAG97991 to AAC98763 encode the human colon cancer associated proteins,
XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX CC human colon cancer antigens can have cytostatic, cardiostatic, muscular,
XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX CC vulnary, nephrotoxic, antineoplastic, antibacterial activities, and
XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,
XX CC proteins and antibodies to the proteins are useful for the prevention,
XX CC treatment and diagnosis of colon disorders, such as colon cancer. The
XX CC polynucleotides may be used in diagnostics and research, such as for
XX CC chromosome identification, and as hybridisation probes. The proteins may
XX CC also be used to prevent diseases such as neural disorders, immune system
XX CC disorders, muscular disorders, reproductive disorders, gastrointestinal
XX CC disorders, wounds, renal disorders, infectious diseases, and
XX CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
XX CC sequences used in the exemplification of the present invention
XX SQ
XX Sequence 91 AA;
XX Query Match 5.2%; Score 194.5; DB 3; Length 91;
XX Best Local Similarity 59.2%; Pred. No. 4.1e-13;
XX Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;
QY 284 PRISPPAGRP-----PPSTRXXRAGGVRRAPGPGSAECSSMETGRGKXPL 335
DB 19 PRISDSRSPACRRRAPPSTRPARACGVRRRPAFGSGAECSSMETGGMKG--- 75
QY 336 ARHAPHYRARA 346
DB 76 RLDDPSIRTRA 86

```

```

RESULT 6
ID AAB53743 standard; protein; 91 AA.
AC AAB53743;
DT 09-MAR-2001 (first entry)
XX

```

```

RESULT 7
ID AAB53817 standard; protein; 73 AA.
AC AAB53817;
DT 09-MAR-2001 (first entry)
XX

```

```

DE Human colon cancer antigen protein sequence SEQ ID NO:1357.

```

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX identification; cytostatic; cardiostatic; neuroprotective; vlnnary;
 KM immunomodulatory; muscular; gynaecological; gastrointestinal;
 KM nephrotropic; antinefactive; antibacterial; gene therapy; wound;
 KM neural disorder; immune system disorder; muscular disorder;
 KM reproductive disorder; gastrointestinal disorder; renal disorder;
 KM infectious disease; cardiovascular disorder.
 XX
 OS Homo sapiens.
 XX
 PM WO200055351-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US005863.
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM;
 XX WPI; 2000-587534/55.
 DR N-PSDB; AAC98574.
 XX
 XX Coln cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer.
 XX
 Claim 11; Page 1932-1933; 2104p; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB5334 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardiostatic, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vlnnary, nephrotropic, antinefactive and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins may
 CC also be used to prevent diseases such as neural disorders, immune system
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal
 CC disorders, wounds, renal disorders, infectious diseases, and
 CC cardiovascular disorders. AAC9764 to AAC98772 and AAB54007 represent
 CC sequences used in the exemplification of the present invention
 CC
 XX
 SQ Sequence 73 AA;
 XX
 Query Match 4.9%; Score 185; DB 3; Length 73;
 Best Local Similarity 69.1%; Pred. No. 3.5e-12;
 Matches 38; Conservative 2; Mismatches 7; Indels 8; Gaps 1
 XX
 QY 284 PRLSPAGP-----PSTRXXKAGRVRRRARGPSSXAECSWSWTGGRK 330
 ||| : |||
 ID 19 PRLADSRSPACPRRRAPXSTRKAAAGRVRRRARGPSSGAECSWSWTGGRK 73
 XX
 RESULT 8
 AAB53266
 ID AAB53266 standard; protein; 7052 AA.
 XX
 AAB53266;
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Human P450RAI-3 protein #3.
 XX
 KM Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
 KM P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
 enzyme.
 XX

```

XX Homo sapiens.
FN MO000295034-A2.
FD 28-NOV-2002.
PF 23-MAY-2002; 2002MO-CAO00758.
PR 23-MAY-2001; 2001US-0292531P.
XX (CYTO-) CYTOCHROMA INC.
PA Wisniewski J, Peckovich PM, Ramsdew H;
PI WPI; 2003-112152/10.
DR N-PSDB; AAD53861.
XX New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT diagnosing, preventing, or treating disorders with aberrant expression or
PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT psoriasis.
PS Example 1; Fig 1; 231pp; English.
XX The invention relates to retinoic acid metabolising cytochrome P450.
CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC invention are useful for diagnosing, preventing, ameliorating and/or
CC treating disorders associated with the aberrant expression or activity of
CC the P450RAI such as diseases related to vitamin A and retinoic acid
CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC present sequence is human P450RAI protein
XX SQ Sequence 7052 AA;
Query Match 4.0%; Score 149; DB 6; Length 7052;
Best Local Similarity 54.2%; Pred. No. 9.7e-06;
Matches 39; Conservative 4; Mismatches 25; Indels 4; Gaps 3
QY 713 DQPOHGETSLQKNTKTXPVLAHTGSLSTSEGGXIXIDAAQLEEAANAVERDXAIAPQ 772
Db 427 DQPOHGETSPILRYNTR-NPGWACACANASYSGGRRT--ATWEVEVAISR-DHATAPQ 482
QY 773 GXERETLSQKKX 784
Db 483 GGQSETPSQKKX 494
RESULT 9
ID AAB53774 standard; protein; 54 AA.
AC AAB53774;
DT 09-MAR-2001 (first entry)
DE Human colon cancer antigen protein sequence SEQ ID NO:1314.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardiotoxic; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW neoplastic; anti-infective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
OS Homo sapiens.
PN WO200055351-A1.
PD 21-SEP-2000.
XX 08-MAR-2000; 2000MO-US005883.

```

XX 12-MAR-1999; 99US-0124270P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI; 2000-597534/55.
XX N-PSDB; AAC98531.
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer.
XX Claim 11; Page 1896; 21049P; English.
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardiocactive, muscular,
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerability, nephrotoxic, anti-infective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridization probes. The proteins may
CC also be used to prevent diseases such as neural disorders, immune system
CC disorders, muscular disorders, reproductive disorders, gastrointestinal
CC disorders, wounds, renal disorders, infectious diseases, and
CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
CC sequences used in the exemplification of the present invention
XX Sequence 54 AA;
SQ
Query Match 3.9%; Score 145; DB 3; Length 54;
Best Local Similarity 96.3%; Pred. No. 6.3e-08;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 205 TTVQVSEFTLXMCCHGNPAQYERNR 231
Db 1 TTVQVSEFTLXMCCHGNPAQYERNR 27
RESULT 10
AAM96009 standard; protein; 54 AA.
XX AAM96009;
AC AAM96009;
XX 21-NOV-2001 (first entry)
DT
XX
XX Human reproductive system related antigen SEQ ID NO: 4667.
DE
XX Human reproductive system related antigen; reproductive system disorder;
KW Cancer; gene therapy.
XX Homo sapiens.
OS
XX W0200155320-A2.
FN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001339.
PF
XX 31-JAN-2000; 2000US-01790652.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR

PR 28-JUN-2000; 2000US-0214686P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216680P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218390P.
PR 26-JUL-2000; 2000US-0220363P.
PR 26-JUL-2000; 2000US-0220364P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229517P.
PR 06-SEP-2000; 2000US-0230337P.
PR 06-SEP-2000; 2000US-0230338P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235884P.
PR 27-SEP-2000; 2000US-0235884P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

QY 205 TVRQVSPFTLLXMCCHGNPAQYERNR 231
 Db 1 TVRQVSPFTLLXMCCHGNPAQYERXR 27

RESULT 12

ID ADC95956 standard; protein; 123 AA.

AC ADC95956;

DT 01-JAN-2004 (first entry)

DE E. faecium protein sequence SEQ ID 5583.

KM Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 abdominal-pelvic infection.

OS Enterococcus faecium.

PN US6583275-B1.

PD 24-JUN-2003.

PF 30-JUN-1998; 98US-00107532.

PR 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2003-799836/75.

DR N-PSDB; ADC92302.

PT New isolated nucleic acid derived from Enterococcus faecium encoding an
 Enterococcus faecium polypeptide useful for detection, prevention and
 treatment of a pathological condition resulting from a bacterial
 infection.

PS Example 1; SEQ ID NO 5583; 243bp; English.

XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to a
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids is useful for diagnosing pathological conditions
 CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of Candida albicans -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.

XX Sequence 123 AA;

Query Match 3.3%; Score 122; DB 7; Length 123;
 Best Local Similarity 35.2%; Pred. No. 6e-05;
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPYKRAKARLLIFSTNTDXESGARSXPFGFAXXVKKVTGTGTLMP 164
 Db 8 GELRWGNHSHSVKAEGLTARPTSRAGTKGLSDPVVPHGRALAQRIKATLGTGLSP 67
 QY 165 SXHSDVAFKSPDVGSYHXEAFTKRWIVHPLIGXSMVXTVVRQ 209

Db 68 RVHIDGEVWHLIDVSSHPGAVVGPCKMAVRPLRRHSMQWNTVVRQ 112

RESULT 13

ID ADC95957 standard; protein; 123 AA.

AC ADC95957;

DT 01-JAN-2004 (first entry)

DE E. faecium protein sequence SEQ ID 5584.

KM Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 abdominal-pelvic infection.

OS Enterococcus faecium.

PN US6583275-B1.

PD 24-JUN-2003.

PF 30-JUN-1998; 98US-00107532.

PR 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2003-799836/75.

DR N-PSDB; ADC92303.

PT New isolated nucleic acid derived from Enterococcus faecium encoding an
 Enterococcus faecium polypeptide useful for detection, prevention and
 treatment of a pathological condition resulting from a bacterial
 infection.

PS Example 1; SEQ ID NO 5584; 243bp; English.

XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to a
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids is useful for diagnosing pathological conditions
 CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of Candida albicans -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.

XX Sequence 123 AA;

Query Match 3.3%; Score 122; DB 7; Length 123;
 Best Local Similarity 35.2%; Pred. No. 6e-05;
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPYKRAKARLLIFSTNTDXESGARSXPFGFAXXVKKVTGTGTLMP 164
 Db 8 GELRWGNHSHSVKAEGLTARPTSRAGTKGLSDPVVPHGRALAQRIKATLGTGLSP 67
 QY 165 SXHSDVAFKSPDVGSYHXEAFTKRWIVHPLIGXSMVXTVVRQ 209

Db 68 RVHIDGEYWHLDVGSSHPGAVVGPKGMAVRPLKPHASWQNVYEQ 112

RESULT 14
AAM42248
ID AAM42248 standard; protein: 96 AA.
XX AAM42248;
AC
XX
XX 19-OCT-2001 (first entry)
DT
XX
DE Human breast or ovarian antigen SEQ ID NO: 125.
XX
XX Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy.
XX Homo sapiens.
XX WO20015324-A2.
XX
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001344.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0188874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0230963P.
PR 26-JUL-2000; 2000US-0230964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0226279P.
PR 18-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227003P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229309P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230438P.
PR 06-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249248P.
PR 17-NOV-2000; 2000US-0249249P.
PR 17-NOV-2000; 2000US-0249250P.

[illegible]

PR 02-OCT-2000; 2000US-0237033P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0253678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-465570/50.
 DR N-PSDB; AAL00321.
 XX
 XX Isolated nucleic acid molecule encoding a reproductive system antigen is
 PT used in preventing, treating or ameliorating a medical condition.
 XX
 PS Claim 11; SEQ ID NO 3009; 1297bp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention
 XX
 SQ Sequence 96 AA;
 Query Match 3.2%; Score 120; DB 4; Length 96;
 Best Local Similarity 46.5%; Pred. No. 7.4e-05;
 Matches 33; Conservative 4; Mismatches 30; Indels 4; Gaps 2;
 Oy 660 SPFLIYVYITNTGLSEQLIHKRXSGTGEKXLIPIVILALXXAKAGSLERVDOPXHG 719
 Db 28 SLFSFVAVAKSTCL--WLIRKTVVPAGHQAQ-PVLPALWEAKAGSLERXKRDOPQHG 83
 Oy 720 ETLSLQKNTKI 730
 Db 84 ETQXLPKIQKL 94

Search completed: April 27, 2004, 10:52:58
 Job time : 55.1947 secs

Query Match 11.6%; Score 435; DB 4; Length 538;
 Best Local Similarity 28.8%; Pred. No. 6.9e-40;
 Matches 182; Conservative 50; Mismatches 216; Indels 184; Gaps 30;

123 LILISTNDXESGASRFXFGFXAXVKTGTGTGLMRPSXSDVAFXSFDVSSYH 182
 1 MILISTNDRESGAS-----DLTFWVLSRQC-----KSYH 33
 183 XEAEFTKWTIV-----HPLIGXSWXTVV-----RQVS-----211
 34 RD-----NMLVAARSKRRRFLILRCFLFSLGSRHQALDCSPTRRRELGLDRRTGL 88
 212 FTLLMCCCHGNPAQYERNRKHLYVVLGAGNAXKXSVGLXNLNASESPXGTIRQ 271
 89 FTLLMCCCHGNPAQYERNRKHLYVVLGAGNAXKXSVGLXNLNASESPXGTIRQ 148
 272 RGASVGLGXPRLSPPAGRPSPSTRXKXAGRVPRAPGPGSXAECSSWETGRKRG 331
 149 RGASVGLGWPPRLSPAGAPPP-----RAP-----RABEG 181
 332 GXPLARH-----APHVRAAEFXXSSTINRHTSACIFMXXXILFLWVDIOXWDCX 382
 182 ACFAARRDRGVRSAALRPGKGAAGAAPSVTHTFTVLVNSARVAPFTIDIOV-HVS 240
 383 XTWXYFYWF-----IEKSYXXXXXRLYKFTSLPSDFPKXERWVRNAPHXYP-PFTYL 434
 241 LLYNEFFSGEISSGSIASPGSSISGLLR--NLHTDFH-----RGCTNLHPYQVIFL 291
 435 LQDFKGYRYQVSKXQIEYXNFCIRGT--DFXIOSCMNDKSRDLOSXNWSOMXY 491
 292 NMKEMSEMP--ISIPLLPLYCRMTSRGTGYKFTYDKNLNIEISALEQIIGFK-----345
 492 ISSSTXSGST-----ELAKSSLI--PTY-----XXXXKGFISNLIHG 528
 346 VVTRTSLAGTCAGAGINLRNTFLVVPAYILLNMLCDHLYLILKTNKGLYQIYFYK 405
 529 IKIKXXVXKLSLAFXQNIKXXPSIGHLV-----CTRHCVCHXSKMFSXGSG 576
 406 KKSNYRKSCLLHFNFKTKNSYLLG-IYTVDPVFTVFOKCS-HGVNANNV-GW-----458
 577 XFCRVXSLTVVLFKSKNLXTXKXFNLRKVSNRTRTXXXITLOISPYHTAS--TCACXL 634
 459 -----RNSLTVRLF--RNETSEPM-----SSTCAKLVTLGGLPEPSHSRSLHTTLAH 506
 635 IPGSCYF-PFYFLSLXKTTFFSPHFSEFLIV 665
 507 VPVILFLAPLFLPPSSSHNPPFPFL-FLFLI 537

RESULT 2
 US-09-643-476-2
 / Sequence 2, Application US/09643476
 / Patent No. 6599495
 / GENERAL INFORMATION:
 / APPLICANT: Evin, Jr., Paul R.
 / TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE
 / OF MAMMASTATIN AND METHODS OF USE
 / NUMBER OF SEQUENCES: 8
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Merchant & Gould P.C.
 / STREET: P.O. Box 2903
 / CITY: Minneapolis
 / STATE: MN
 / COUNTRY: USA
 / ZIP: 55402-0903
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Diskette
 / OPERATING SYSTEM: DOS
 / SOFTWARE: FASTSEQ for Windows Version 2.0
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/09/643,476
 / FILING DATE: 22-Aug-2000

CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/027,315
 FILING DATE: 03-OCT-1996
 APPLICATION NUMBER: PCT/US97/18026
 FILING DATE: 03-OCT-1997
 APPLICATION NUMBER: 08/943,828
 FILING DATE: 03-OCT-1997
 APPLICATION NUMBER: 08/994,076
 FILING DATE: 19-DEC-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Weaver, Katie G.
 REGISTRATION NUMBER: 43,245
 REFERENCE/DOCKET NUMBER: 4273.1USC2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEFAX: 612-332-9081
 TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 538 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-643-476-2

Query Match 11.6%; Score 435; DB 4; Length 538;
 Best Local Similarity 28.8%; Pred. No. 6.9e-40;
 Matches 182; Conservative 50; Mismatches 216; Indels 184; Gaps 30;

123 LILISTNDXESGASRFXFGFXAXVKTGTGTGLMRPSXSDVAFXSFDVSSYH 182
 1 MILISTNDRESGAS-----DLTFWVLSRQC-----KSYH 33
 183 XEAEFTKWTIV-----HPLIGXSWXTVV-----RQVS-----211
 34 RD-----NMLVAARSKRRRFLILRCFLFSLGSRHQALDCSPTRRRELGLDRRTGL 88
 212 FTLLMCCCHGNPAQYERNRKHLYVVLGAGNAXKXSVGLXNLNASESPXGTIRQ 271
 89 FTLLMCCCHGNPAQYERNRKHLYVVLGAGNAXKXSVGLXNLNASESPXGTIRQ 148
 272 RGASVGLGXPRLSPPAGRPSPSTRXKXAGRVPRAPGPGSXAECSSWETGRKRG 331
 149 RGASVGLGWPPRLSPAGAPPP-----RAP-----RABEG 181
 332 GXPLARH-----APHVRAAEFXXSSTINRHTSACIFMXXXILFLWVDIOXWDCX 382
 182 ACFAARRDRGVRSAALRPGKGAAGAAPSVTHTFTVLVNSARVAPFTIDIOV-HVS 240
 383 XTWXYFYWF-----IEKSYXXXXXRLYKFTSLPSDFPKXERWVRNAPHXYP-PFTYL 434
 241 LLYNEFFSGEISSGSIASPGSSISGLLR--NLHTDFH-----RGCTNLHPYQVIFL 291
 435 LQDFKGYRYQVSKXQIEYXNFCIRGT--DFXIOSCMNDKSRDLOSXNWSOMXY 491
 292 NMKEMSEMP--ISIPLLPLYCRMTSRGTGYKFTYDKNLNIEISALEQIIGFK-----345
 492 ISSSTXSGST-----ELAKSSLI--PTY-----XXXXKGFISNLIHG 528
 346 VVTRTSLAGTCAGAGINLRNTFLVVPAYILLNMLCDHLYLILKTNKGLYQIYFYK 405
 529 IKIKXXVXKLSLAFXQNIKXXPSIGHLV-----CTRHCVCHXSKMFSXGSG 576
 406 KKSNYRKSCLLHFNFKTKNSYLLG-IYTVDPVFTVFOKCS-HGVNANNV-GW-----458
 577 XFCRVXSLTVVLFKSKNLXTXKXFNLRKVSNRTRTXXXITLOISPYHTAS--TCACXL 634
 459 -----RNSLTVRLF--RNETSEPM-----SSTCAKLVTLGGLPEPSHSRSLHTTLAH 506

Wed Apr 28 09:57:47 2004

us-10-028-952a-10.rat

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:51:05 / Search time 19.8859 Seconds
(without alignments)
2035.354 Million cell updates/sec

Title: US-10-028-952A-10

Perfect score: 3752
Sequence: 1 IRHGHGRTYEVNKKKAPG.....DKAIALQPGKRETTLSQKK 784

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:
1: /cgn2.6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2.6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2.6/prodata/2/1aa/5A_COMB.pep.*
4: /cgn2.6/prodata/2/1aa/5B_COMB.pep.*
5: /cgn2.6/prodata/2/1aa/5A_COMB.pep.*
6: /cgn2.6/prodata/2/1aa/5B_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	435	11.6	538	US-08-994-076-2
2	435	11.6	538	US-09-643-476-2
3	133.5	3.6	497	US-09-252-991A-22620
4	122.5	3.3	67	US-09-621-976-3994
5	122	3.3	123	US-09-107-532A-5583
6	122	3.3	123	US-09-107-532A-5584
7	122	3.3	123	US-09-134-000C-6555
8	120.5	3.2	555	US-09-252-991A-28848
9	119	3.2	97	US-09-134-001C-3072
10	119	3.2	97	US-09-134-001C-3726
11	119	3.2	97	US-09-134-001C-3875
12	119	3.2	97	US-09-134-001C-4192
13	119	3.2	97	US-09-134-001C-4844
14	117	3.1	106	US-09-134-001C-3312
15	116.5	3.1	456	US-09-252-991A-17335
16	116	3.1	82	US-09-134-000C-4749
17	115.5	3.1	653	US-09-252-991A-21591
18	115	3.1	405	US-09-252-991A-24483
19	115	3.1	425	US-09-252-991A-20467
20	114.5	3.1	728	US-09-252-991A-22613
21	113.5	3.0	719	US-09-252-991A-23660
22	113	3.0	568	US-09-252-991A-23264
23	112.5	3.0	390	US-09-252-991A-33044
24	111	3.0	397	US-09-252-991A-24025
25	110.5	2.9	1020	US-09-252-991A-28870
26	110	2.9	348	US-09-252-991A-25605
27	109.5	2.9	1113	US-09-252-991A-25215

28	108.5	2.9	312	US-09-252-991A-20216	Sequence 20216, A
29	108	2.9	274	US-09-252-991A-28347	Sequence 28347, A
30	108	2.9	323	US-09-252-991A-23885	Sequence 23885, A
31	108	2.9	496	US-09-252-991A-22794	Sequence 22794, A
32	107	2.9	472	US-09-252-991A-25386	Sequence 25386, A
33	106.5	2.8	309	US-09-252-991A-25386	Sequence 25386, A
34	106.5	2.8	472	US-09-252-991A-31978	Sequence 31978, A
35	106	2.8	70	US-09-621-976-4551	Sequence 4551, AP
36	106	2.8	530	US-09-252-991A-20846	Sequence 20846, A
37	105.5	2.8	485	US-09-252-991A-22450	Sequence 22450, A
38	105	2.8	260	US-09-252-991A-28669	Sequence 28669, A
39	105	2.8	628	US-09-252-991A-32727	Sequence 32727, A
40	104.5	2.8	478	US-09-252-991A-24169	Sequence 24169, A
41	104	2.8	693	US-09-252-991A-26071	Sequence 26071, A
42	103	2.7	603	US-09-252-991A-22674	Sequence 22674, A
43	102.5	2.7	354	US-09-252-991A-26135	Sequence 26135, A
44	102.5	2.7	210	US-09-252-991A-19986	Sequence 19986, A
45	102.5	2.7	606	US-09-252-991A-18233	Sequence 18233, A

ALIGNMENTS

RESULT 1

US-08-994-076-2
Sequence 2, Application US/08994076

Patent No. 650937

GENERAL INFORMATION:
APPLICANT: Evin, Jr., Paul R.
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE
TITLE OF INVENTION: OF MAMMASTATIN AND METHODS OF USE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:
ADDRESS: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 6500937west Center, 90 South 7th Street

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,076

FILING DATE: 19-DEC-1997

CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/027,315

FILING DATE: 03-OCT-1996

APPLICATION NUMBER: PCT/US97/18026

FILING DATE: 03-OCT-1997

APPLICATION NUMBER: 08/943,828

FILING DATE: 03-OCT-1997

ATTORNEY/AGENT INFORMATION:
NAME: Kettelberger, Denise X

REGISTRATION NUMBER: 33,924

REFERENCE/DOCKET NUMBER: 4273, 1US11

TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-994-076-2

QY 635 IPGSCVF-PFYFLSLXITTPESHPFSEFLIV 665
DB 507 VVILLFLAPXLFPPSSHNPFPPFL-FLFLP 537

RESULT 3

US-09-252-991A-23620
Sequence 23620, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Mate J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,768
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23620
LENGTH: 497
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23620

Query Match 3.6%; Score 133.5; DB 4; Length 497;
Best Local Similarity 22.7%; Pred. No. 3.3e-06;
Matches 97; Conservative 30; Mismatches 135; Indels 165; Gaps 20;

QY 18 AGGAPPVSPRGA-----RGXRRPCCPVYKHYSDXXTPVRRGGER-- 61
DB 112 APAAPARRRALLAQFRSQVRPATAREGGRSPDPER-----RLRRGPLRGGHPAFR 165
QY 62 ----GALAXGAKRPA-----RRPGATRSQXSARMGVXLGRYTCQTXOV-----SXGE 106
DB 166 RAIPGAAAPGRPRPALPFRYRPGRRRAGAPR---RMGQRLRAGALLAAGAAGATGA 222
QY 107 IREDN-----LPMXR-----RAKRLILITSTVDXSG--- 136
DB 223 VRTTRPGVGRPRRLRTARPGOPPGPLPARQGRPGRAGSDLRAPRATAGRSAGDRQ 282
QY 137 ---ASRSXP-----GFXAXXVKTGTG-----LWPSXSHSVAFXSFDVGSY 181
DB 283 GGRRLPAGPGLPQRTGILHARRORRVADGPGSVRAPAMGGHDL----- 331
QY 182 HXAEFTKRWIVHPLIGXXSWYTVRQVSTLLXMCCHGNPAQYERNRXLVYL 241
DB 332 --PGQPQARQLAEPGTG-----PAPARQSRRLRHLRLH 364
QY 242 XGANAAXXSVGLXNASKSESRPGTTRQR-----GAS-----VGLGKXPRLSPPA 290
DB 365 RPAEGGQHPRG---AGRAPMDAGHLPRRRRADAEGAGELRRVG---VLLAA 415
QY 291 GPPPTSTXKXAGVPRRAPGPGSXAECPSMETGRGKGPXARAPHYRABAEFXX 350
DB 416 GHRLP-----AGRRPRAPRSGA-----PGGTGAAYGDHALARA--AA 453
QY 351 SSTINR 357
DB 454 AATVHR 460

RESULT 4
US-09-621-976-3994
Sequence 3994, Application US/09621976
Patent No. 6639083
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Joberc, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTE and Encoded Human Proteins.

FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 3994
LENGTH: 67
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -26..-1
US-09-621-976-3994

Query Match 3.3%; Score 122.5; DB 4; Length 67;
Best Local Similarity 52.5%; Pred. No. 3.8e-06;
Matches 31; Conservative 5; Mismatches 22; Indels 1; Gaps 1;

QY 726 KTKIXXVLAHTCSLSYSGXGXIDXAQVEVAAVRYDXATALPGXERETLSQKK 784
DB 8 ENTXMLGWVCAQSPSYSGRGRRIAWTREVAVSR-DQATALPGQORETLSQKK 65

RESULT 5

US-09-107-532A-5583
Sequence 5583, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucelte-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Denike
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5583:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...123
SEQUENCE DESCRIPTION: SEQ ID NO: 5583:
US-09-107-532A-5583

Query Match 3.3%; Score 122; DB 4; Length 123;
 Best Local Similarity 35.2%; Pred. No. 9.8e-06;
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDRLPMWRARAKRLILFSTNTDXESGASRSPFGFAXXVRYVTGITGLMRP 164
 DB 8 GSLRWGNHNSKSVKAEGLSLTARPTSRAGTKVGLSDPVPHGRALAKRIKATIGITGLSP 67
 QY 165 SXHSDVAFKSPDVGSSTYHKEAFTKRWIVHPLIGXSWXTYVRQ 209
 DB 68 RVHIDGEVWHLDVGSSHPGAVVGPKGMAVRFKRAHSAWQNVVRQ 112

RESULT 6
 US-09-107-532A-5584
 Sequence 5584, Application US/09107532A
 Patent No. 6583275

GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02154

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051,571
 FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
 NAME: Ariniello, Pamela Deneka
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5584:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 123 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium
 FEATURES:
 NAME/KEY: misc_feature
 LOCATION: (B) LOCATION 1...123
 SEQUENCE DESCRIPTION: SEQ ID NO: 5584:
 US-09-107-532A-5584

Query Match 3.3%; Score 122; DB 4; Length 123;
 Best Local Similarity 35.2%; Pred. No. 9.8e-06;
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDRLPMWRARAKRLILFSTNTDXESGASRSPFGFAXXVRYVTGITGLMRP 164
 DB 8 GSLRWGNHNSKSVKAEGLSLTARPTSRAGTKVGLSDPVPHGRALAKRIKATIGITGLSP 67

QY 165 SXHSDVAFKSPDVGSSTYHKEAFTKRWIVHPLIGXSWXTYVRQ 209
 DB 68 RVHIDGEVWHLDVGSSHPGAVVGPKGMAVRFKRAHSAWQNVVRQ 112

RESULT 7
 US-09-134-000C-6555
 Sequence 6555, Application US/09134000C
 Patent No. 6617156

GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 032796-032
 CURRENT APPLICATION NUMBER: US/09/134,000C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/055,778
 PRIOR FILING DATE: 1997-08-15
 NUMBER OF SEQ ID NOS: 6812
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 6555

TYPE: PRT

ORGANISM: Enterococcus faecalis
 US-09-134-000C-6555

Query Match 3.3%; Score 122; DB 4; Length 123;
 Best Local Similarity 35.2%; Pred. No. 9.8e-06;
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDRLPMWRARAKRLILFSTNTDXESGASRSPFGFAXXVRYVTGITGLMRP 164
 DB 8 GSLRWGNHNSKSVKAEGLSLTARPTSRAGTKVGLSDPVPHGRALAKRIKATIGITGLSP 67
 QY 165 SXHSDVAFKSPDVGSSTYHKEAFTKRWIVHPLIGXSWXTYVRQ 209
 DB 68 RVHIDGEVWHLDVGSSHPGAVVGPKGMAVRFKRAHSAWQNVVRQ 112

RESULT 8
 US-09-252-991A-28848
 Sequence 28848, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196-136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 28848
 LENGTH: 555
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-28848

Query Match 3.2%; Score 120.5; DB 4; Length 555;
 Best Local Similarity 22.5%; Pred. No. 0.00011;
 Matches 93; Conservative 21; Mismatches 144; Indels 155; Gaps 19;

QY 15 KKEAPGAPVSP-----RGARGGKRRPGPPVTKYHSDRXXTDVR-RGGEPRGALA 65
 DB 83 REHCPHDPAAKPDHPPRRWRGLRGGRRP--PTDRLEAAGBEERHPRRLRSAPAGARC 140
 QY 66 XG-----AKRPARP--GATSGSARWGVXLGRTQOTXXQVSG-----ELR 108
 DB 141 FGPRORPGKPDADADAPRRDVRGVRAGQ---GRAHGHCRRLRRRVAGGAPLEAR 136

```
QY 109 EDENLFWXRRAK---ARLLIFSTNTDSESGASRFXPGFXAXYV-----RKVT 155
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
Db 197 PHRAAPDGLPRNRHAGTVGLDHHRRPRAGSRVLMCGMAAPVOAARAPDRFNRR 256
; US-09-134-001C-3726
QY 156 TGITGLMRPSXHSVAKSPVGVSSYHXEAFTRMIVHPLIGXSWXVTVVQVSTLL 215
; Query Match
; Best Local Similarity 37.8%; Pred. No. 1.6e-05; Length 97;
; Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;
Db 257 LARPAALRRAGHLDPAG-----RRR----- 275
; US-09-134-001C-3072
QY 216 XMCCCHGPAQYEENRERXRLVYVLGXGANGAKXXSVGLXLMNKSSESPXGTLRORGA 275
; Sequence 3072, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3072
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3072
Query Match
; Best Local Similarity 3.2%; Score 119; DB 4; Length 97;
; Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;
QY 128 STNTDSESGASRFXPGFXAXXVRKVTGTTGLMRPSXHSVDVAFSPDVGSYHXAEF 187
; Db
; 8 TSRAQSKDGLSDPVVPHGRALAQRIKATPGITGLSPRVHIDGEVWHLDVGSSHPGAVVG 67
; US-09-134-001C-3726
QY 188 TKRWIVHPLIGXSWXVTVVQ 209
; Db
; 68 PKGMAVRLPKRYASWVQNVVRQ 89
; US-09-134-001C-3726
; Sequence 3726, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3726
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
```

```
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3726
Query Match
; Best Local Similarity 3.2%; Score 119; DB 4; Length 97;
; Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;
QY 128 STNTDSESGASRFXPGFXAXXVRKVTGTTGLMRPSXHSVDVAFSPDVGSYHXAEF 187
; Db
; 8 TSRAQSKDGLSDPVVPHGRALAQRIKATPGITGLSPRVHIDGEVWHLDVGSSHPGAVVG 67
; US-09-134-001C-3726
QY 188 TKRWIVHPLIGXSWXVTVVQ 209
; Db
; 68 PKGMAVRLPKRYASWVQNVVRQ 89
; US-09-134-001C-3875
; Sequence 3875, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3875
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3875
Query Match
; Best Local Similarity 37.8%; Pred. No. 1.6e-05; Length 97;
; Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;
QY 128 STNTDSESGASRFXPGFXAXXVRKVTGTTGLMRPSXHSVDVAFSPDVGSYHXAEF 187
; Db
; 8 TSRAQSKDGLSDPVVPHGRALAQRIKATPGITGLSPRVHIDGEVWHLDVGSSHPGAVVG 67
; US-09-134-001C-4192
; Sequence 4192, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4192
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
```



```

OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (249)..(250)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (255)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (266)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (281)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (283)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (299)..(300)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (316)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (333)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (349)..(350)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (366)..(368)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (378)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (382)..(383)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (386)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (398)..(401)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (416)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (427)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (433)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (449)..(450)
OTHER INFORMATION: Xaa is any amino acid or may be absent

```

```

FEATURE:
NAME/KEY: UNSURE
LOCATION: (456)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (466)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (483)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (499)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

Query Match 94.5%; Score 3546; DB 14; Length 784;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 IRHEGEETXEVXNCKEAPGAPVPSPRGARGXRRPCGPVYKTHYSDKXTDPVRRGCEP 60
DB 1 IRHEGEETXEVXNCKEAPGAPVPSPRGARGXRRPCGPVYKTHYSDKXTDPVRRGCEP 60
QY 61 RGALAXGAKRPAARRPGATRSXGASARWGVXIGRYTCOTXOVXSGELREDENTLPMXBRAX 120
DB 61 RGALAXGAKRPAARRPGATRSXGASARWGVXIGRYTCOTXOVXSGELREDENTLPMXBRAX 120
QY 61 RGALAXGAKRPAARRPGATRSXGASARWGVXIGRYTCOTXOVXSGELREDENTLPMXBRAX 120
DB 61 RGALAXGAKRPAARRPGATRSXGASARWGVXIGRYTCOTXOVXSGELREDENTLPMXBRAX 120
QY 121 ARLLIFSTNTDXESGASRSFPFGFXAXXVRYKVTGTITGLMRPSXHSDDVAFXSFDVGS 180
DB 121 ARLLIFSTNTDXESGASRSFPFGFXAXXVRYKVTGTITGLMRPSXHSDDVAFXSFDVGS 180
QY 121 ARLLIFSTNTDXESGASRSFPFGFXAXXVRYKVTGTITGLMRPSXHSDDVAFXSFDVGS 180
DB 121 ARLLIFSTNTDXESGASRSFPFGFXAXXVRYKVTGTITGLMRPSXHSDDVAFXSFDVGS 180
QY 181 YHKEEFTKRWTVHPLIGXSMVTVVROVSFTLLMCCGNGNPAQYERRRXXHLYVL 240
DB 181 YHKEEFTKRWTVHPLIGXSMVTVVROVSFTLLMCCGNGNPAQYERRRXXHLYVL 240
QY 181 YHKEEFTKRWTVHPLIGXSMVTVVROVSFTLLMCCGNGNPAQYERRRXXHLYVL 240
DB 181 YHKEEFTKRWTVHPLIGXSMVTVVROVSFTLLMCCGNGNPAQYERRRXXHLYVL 240
QY 241 GXGANGAKXXSVGLXLNASKESREPGTIRORRASVGLGXPRLSPPAGRPSTTRXX 300
DB 241 GXGANGAKXXSVGLXLNASKESREPGTIRORRASVGLGXPRLSPPAGRPSTTRXX 300
QY 241 GXGANGAKXXSVGLXLNASKESREPGTIRORRASVGLGXPRLSPPAGRPSTTRXX 300
DB 241 GXGANGAKXXSVGLXLNASKESREPGTIRORRASVGLGXPRLSPPAGRPSTTRXX 300
QY 301 RAGGVVPRRARPQSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSTIHNRHTS 360
DB 301 RAGGVVPRRARPQSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSTIHNRHTS 360
QY 301 RAGGVVPRRARPQSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSTIHNRHTS 360
DB 301 RAGGVVPRRARPQSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSTIHNRHTS 360
QY 361 ACIFPKXXILFLMVDIOXMDCKXTWXXFYFMEIEKSSYXXXXRLYKFTSLPDPFKXBRMV 420
DB 361 ACIFPKXXILFLMVDIOXMDCKXTWXXFYFMEIEKSSYXXXXRLYKFTSLPDPFKXBRMV 420
QY 361 ACIFPKXXILFLMVDIOXMDCKXTWXXFYFMEIEKSSYXXXXRLYKFTSLPDPFKXBRMV 420
DB 361 ACIFPKXXILFLMVDIOXMDCKXTWXXFYFMEIEKSSYXXXXRLYKFTSLPDPFKXBRMV 420
QY 421 WRNAPHXYPPTXLIQNDFKGRYIQQVXXRQIERYNFCIRGTDFTXIQSCNNKDCGRDL 480
DB 421 WRNAPHXYPPTXLIQNDFKGRYIQQVXXRQIERYNFCIRGTDFTXIQSCNNKDCGRDL 480
QY 421 WRNAPHXYPPTXLIQNDFKGRYIQQVXXRQIERYNFCIRGTDFTXIQSCNNKDCGRDL 480
DB 421 WRNAPHXYPPTXLIQNDFKGRYIQQVXXRQIERYNFCIRGTDFTXIQSCNNKDCGRDL 480
QY 481 OSXWKSQMKYISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVXKLSL 540
DB 481 OSXWKSQMKYISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVXKLSL 540
QY 481 OSXWKSQMKYISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVXKLSL 540
DB 481 OSXWKSQMKYISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVXKLSL 540
QY 541 AFEPONIKXPSIGHLYCTRHCVCHXSXMPFWSXCSQXFCRRXSLTVVRLFSKNLXTXX 600
DB 541 AFEPONIKXPSIGHLYCTRHCVCHXSXMPFWSXCSQXFCRRXSLTVVRLFSKNLXTXX 600
QY 541 AFEPONIKXPSIGHLYCTRHCVCHXSXMPFWSXCSQXFCRRXSLTVVRLFSKNLXTXX 600
DB 541 AFEPONIKXPSIGHLYCTRHCVCHXSXMPFWSXCSQXFCRRXSLTVVRLFSKNLXTXX 600
QY 601 FNLKRVSRTRTXITXITLOISPYHTASTCAXLIPSCVYPPFFLSLXTTSPSPHFS 660
DB 601 FNLKRVSRTRTXITXITLOISPYHTASTCAXLIPSCVYPPFFLSLXTTSPSPHFS 660
QY 601 FNLKRVSRTRTXITXITLOISPYHTASTCAXLIPSCVYPPFFLSLXTTSPSPHFS 660
DB 601 FNLKRVSRTRTXITXITLOISPYHTASTCAXLIPSCVYPPFFLSLXTTSPSPHFS 660
QY 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXQGE 720
DB 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXQGE 720
QY 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXQGE 720
DB 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXQGE 720
QY 721 TLSIQKNTKIKXPYLAFHTCSLSYSEGKXXIDXAQVEEAANVXDYALIQPKERETLS 780
DB 721 TLSIQKNTKIKXPYLAFHTCSLSYSEGKXXIDXAQVEEAANVXDYALIQPKERETLS 780

```

```

DB 721 TLSIQKNTKIKXPYLAFHTCSLSYSEGKXXIDXAQVEEAANVXDYALIQPKERETLS 780
QY 781 OKKK 784
DB 781 OKKK 784

RESULT 2
US-10-028-952a-9
Sequence 9, Application US/10028952A
Publication No. US20030157576A1
GENERAL INFORMATION:
APPLICANT: Eryin, Jr., Paul R.
TITLE OR INVENTION: EPITHELIAL CELL GROWTH INHIBITORS
CURRENT APPLICATION NUMBER: US/10/028,952A
CURRENT FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: PCT/US00/16900
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 60/139,995
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 793
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (15)..(17)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (42)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (45)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (49)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (65)..(66)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (82)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (99)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (115)..(116)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (122)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (132)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (136)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

LOCATION: (149)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (165) ..(166)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (174)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (179)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (182)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (199)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (205)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (215) ..(216)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (232)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (236)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (249)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (265) ..(266)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (274)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (282)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (287)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (299)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (313)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (315) ..(316)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (332)

OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (349)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (365) ..(366)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (380)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (382)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (399)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (415) ..(416)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (432)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (449)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (465) ..(466)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (482)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (499)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (515) ..(516)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (532)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (544)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

Query Match 34.3%; Score 1288; DB 14; Length 793;
Best Local Similarity 80.3%; Pred. No. 7.6e-122; Indels 4; Gaps 2;
Matches 282; Conservative 7; Mismatches 58;

Qy 5 HGEETXEVXKXKXAPGAPVSPRGARGXRRPCGPVKYHSDRXXYTDVYRGGEBRGAL 64
Db 37 HGEETXEVXKXKXAPGAPVSPRGARGXRRPCGPVKYHSDRXXYTDVYRGGEBRGAL 96
Qy 65 AXGAKRPAARRPGATNSGXARWGVXIGRTCTQTXKQVSGELREDRNLPEWXRARARLI 124

Db 97 ASAXRPAARFGAARSGXAXAMGYLGRYTCQVYXQVSKSELEDRLLPMSXAKRLLI 156
QY 125 LIFSTNTDXESGASRFXFPFGXAXXVAKVTTGTLGLRPSXSHDVAFXSDVGSYHXE 184
Db 157 LIFSTNTDXSGASRFXFPFGXAXXVAKVTTGTLGLRPSXSHDVAFXSDVGSYHXE 216
QY 185 AEFYKMTIVHPLIGXSWVXTVVRQVSFTLLMCCCHGNPAQYERNRRLVYVLGXGA 244
Db 217 AEFYKMTIVHPLIGXSWVXTVVRQVSFTLLMCCCHGNPAQYERNRRLVYVLGXGA 276
QY 245 NGAKXXSVGLXNLNASKSRSPXGTRORRGASVGLGXSPRLSPAGPSPSTXXRAGG 304
Db 277 NGAKXXSVGLXNLNASKSRSPXGTRORRGASVGLGXSPRLSPAGPSPSTXXRAGG 336
QY 305 RVPRRAPPGSXAECPSSMETGRG-RKGGXPRLARHAPVRAAFEXXSSTI 354
Db 337 RAPRA--GTGVRCXVPFVLNGARPERRPXXRPSRARSMTGWRXTXSI 384

RESULT 3

US-09-925-299-1334
; Sequence 1334, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925, 299

PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883

PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1334

LENGTH: 55
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: SITE
LOCATION: (10)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (49)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (52)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (54)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (55)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1334

Query Match 6.3%; Score 238; DB 9; Length 55;
Best Local Similarity 86.5%; Pred. No. 2.1e-16;
Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 37 CGPVPVYHXSDBRFTDPVRRGEPFGALXGAKRPAARRPGATRSXSARWG 88
Db 2 CGPVPVYHXSDBRFTDPVRRGEPFGALXGAKRPAARRPGATRSXSARWG 53

RESULT 4
US-09-925-299-1334
; Sequence 1334, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925, 299

PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883

PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1334

LENGTH: 55
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: SITE
LOCATION: (10)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (49)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (52)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (54)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
LOCATION: (55)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1334

Query Match 6.3%; Score 238; DB 9; Length 55;
Best Local Similarity 86.5%; Pred. No. 2.1e-16;
Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 37 CGPVPVYHXSDBRFTDPVRRGEPFGALXGAKRPAARRPGATRSXSARWG 88
Db 2 CGPVPVYHXSDBRFTDPVRRGEPFGALXGAKRPAARRPGATRSXSARWG 53

RESULT 5
US-09-925-299-1277
; Sequence 1277, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925, 299

PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883

PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1277

LENGTH: 40
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: SITE
LOCATION: (33)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1277

Query Match 5.3%; Score 199; DB 9; Length 40;
Best Local Similarity 90.0%; Pred. No. 1.3e-12;
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 202 WXTVTVRQVSFTLLMCCCHGNPAQYERNRRLVYVLGXGA 241

Db 1 WYTVVRQVSTFLMCCCHGNPAQYERNR3FXHLVYVLG 40

RESULT 6
US-09-925-299-1277
; Sequence 1277, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1277
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (33)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1277

Query Match 5.3%; Score 199; DB 10; Length 40;
Best Local Similarity 90.0%; Pred. No. 1,3e-12;
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 202 WYTVVRQVSTFLMCCCHGNPAQYERNR3FXHLVYVLG 241
Db 1 WYTVVRQVSTFLMCCCHGNPAQYERNR3FXHLVYVLG 40

RESULT 7
US-09-925-299-1283
; Sequence 1283, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1283
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1283

Query Match 5.2%; Score 194.5; DB 9; Length 91;
Best Local Similarity 59.2%; Pred. No. 1e-11;
Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Qy 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRAAPGSGXACEPSWETGRGKGGXPL 335
Db 1 PRLSPPAGRP-----PPSTRXKXAGRVPRRAAPGSGXACEPSWETGRGKGGXPL 335

Db 19 PRLSDSRSPACPRRRAAPPSTRPARAGRVPRRAAPGSGAECPSWETGGMKGG--- 75

Qy 336 ARAPHVRA 346
Db 76 RLEDPGLRTA 86

RESULT 8
US-09-925-299-1283
; Sequence 1283, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1283
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1283

Query Match 5.2%; Score 194.5; DB 10; Length 91;
Best Local Similarity 59.2%; Pred. No. 1e-11;
Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Qy 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRAAPGSGXACEPSWETGRGKGGXPL 335
Db 19 PRLSDSRSPACPRRRAAPPSTRPARAGRVPRRAAPGSGAECPSWETGGMKGG--- 75

Qy 336 ARAPHVRA 346
Db 76 RLEDPGLRTA 86

RESULT 9
US-09-925-299-1357
; Sequence 1357, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1357
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

Query Match 5.2%; Score 194.5; DB 9; Length 91;
Best Local Similarity 59.2%; Pred. No. 1e-11;
Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Qy 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRAAPGSGXACEPSWETGRGKGGXPL 335
Db 1 PRLSPPAGRP-----PPSTRXKXAGRVPRRAAPGSGXACEPSWETGRGKGGXPL 335

NAME/KEY: SITE
LOCATION: (42)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1357

Query Match 4.9%; Score 185; DB 9; Length 73;
Best Local Similarity 69.1%; Pred. No. 7.2e-11;
Matches 38; Conservative 2; Mismatches 7; Indels 8; Gaps 1;

QY 284 PRLSPPRGP-----PSTRXXRAGGVPRRPAFGSGXAECPSSWETGRGRK 330
DB 19 PRLASDRSPACPRRRAAPKSTRXARAGGVPRRPAFGSGAECPPSSWETGRGRK 73

RESULT 10
US-09-925-299-1357
Sequence 1357, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1357
LENGTH: 73
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (38)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (42)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1357

Query Match 4.9%; Score 185; DB 10; Length 73;
Best Local Similarity 69.1%; Pred. No. 7.2e-11;
Matches 38; Conservative 2; Mismatches 7; Indels 8; Gaps 1;

QY 284 PRLSPPRGP-----PSTRXXRAGGVPRRPAFGSGXAECPSSWETGRGRK 330
DB 19 PRLASDRSPACPRRRAAPKSTRXARAGGVPRRPAFGSGAECPPSSWETGRGRK 73

RESULT 11
US-10-083-357-1137
Sequence 1137, Application US/10083357
Publication No. US20030054370A1
GENERAL INFORMATION:
APPLICANT: Qiangdong Zeng et al.
TITLE OF INVENTION: Systemic Discovery of New Genes
FILE REFERENCE: 032796-090
CURRENT APPLICATION NUMBER: US/10/083,357
CURRENT FILING DATE: 2002-02-27
NUMBER OF SEQ ID NOS: 1346
SEQ ID NO 1137
LENGTH: 57
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1137

Query Match 4.7%; Score 178; DB 14; Length 57;
Best Local Similarity 66.1%; Pred. No. 2.8e-10;
Matches 37; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 117 RRAKARLLIFSTNTDMSGASRSFPGFAXXVRRKTTGTTGIMRPSXHSVDVAF 172
DB 2 KRVKAPLLIFSVNTNHSVAVRSPSPSEFARCARKVTGTTGIMQSSVHSDIAF 57

RESULT 12
US-10-083-357-1147
Sequence 1147, Application US/10083357
Publication No. US20030054370A1
GENERAL INFORMATION:
APPLICANT: Qiangdong Zeng et al.
TITLE OF INVENTION: Systemic Discovery of New Genes
FILE REFERENCE: 032796-090
CURRENT APPLICATION NUMBER: US/10/083,357
CURRENT FILING DATE: 2002-02-27
NUMBER OF SEQ ID NOS: 1346
SEQ ID NO 1147
LENGTH: 57
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1147

Query Match 4.6%; Score 174; DB 14; Length 57;
Best Local Similarity 64.3%; Pred. No. 7.1e-10;
Matches 36; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 117 RRAKARLLIFSTNTDMSGASRSFPGFAXXVRRKTTGTTGIMRPSXHSVDVAF 172
DB 2 KRVKAPLLIFSVNTNHSVAVRSPSPSEFARCARKVTGTTGIMQSSVHSDIAF 57

RESULT 13
US-09-925-299-1314
Sequence 1314, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1314
LENGTH: 54
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-299-1314

Query Match 3.9%; Score 145; DB 9; Length 54;
Best Local Similarity 96.3%; Pred. No. 5.9e-07;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 205 TVRQVSTFLMCCCHGNPAQYERNR 231
DB 1 TVRQVSTFLMCCCHGNPAQYERNR 27

RESULT 14
US-09-925-299-1314
Sequence 1314, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1556
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 1314
 ; LENGTH: 54
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-925-289-1314

Query Match 3.9%; Score 145; DB 10; Length 54;
 Best Local Similarity 96.3%; Pred. No. 5.9e-07;
 Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 205 TVRQVSTFLMCCCHGNPAQYERNR 231
 DB 1 TVRQVSTFLMCCCHGNPAQYERNR 27

RESULT 15
 US-09-764-891-4667
 ; Sequence 4667, Application US//09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 4667
 ; LENGTH: 54
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (44)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-891-4667

Query Match 3.9%; Score 145; DB 10; Length 54;
 Best Local Similarity 96.3%; Pred. No. 5.9e-07;
 Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 205 TVRQVSTFLMCCCHGNPAQYERNR 231
 DB 1 TVRQVSTFLMCCCHGNPAQYERNR 27

Search completed: April 27, 2004, 11:05:14
 Job time : 41.2689 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:46:49 ; Search time 16.903 Seconds
(without alignments)
4461.589 Million cell updates/sec

Title: US-10-028-952A-10
Perfect score: 3752
Sequence: 1 IRHGHGETXEVXNKXKAPG.....DXAIALQPGXRETLISQKXK 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	3.4	613	4 C40201	artifact-warning s
2	111	3.0	673	4 F40201	artifact-warning s
3	104	2.8	660	1 QOBB3	BHFI1 protein - hu
4	103	2.7	376	2 C75580	adenine deaminase-
5	103	2.7	597	4 E40201	artifact-warning s
6	100	2.7	627	4 A40201	artifact-warning s
7	96	2.6	552	2 F75311	ABC transporter, A
8	95	2.5	705	2 A35363	synapsin I splice
9	94	2.5	1460	1 EDBE1F	immediate-early pr
10	93	2.5	450	2 A38316	alpha-2-adrenergic
11	93	2.5	494	2 G84348	hypothetical prote
12	90.5	2.4	579	4 D40201	artifact-warning s
13	90	2.4	529	2 F45134	hypothetical prote
14	89.5	2.4	769	2 T51112	beta-glucosidase (
15	89.5	2.4	1219	2 S54570	probable membrane
16	88.5	2.4	224	2 E72704	hypothetical prote
17	87	2.3	514	1 W2W15	B2 protein - human
18	87	2.3	706	2 E30411	synapsin Ia - Dovi
19	86	2.3	373	2 A47234	homeobox protein H
20	86	2.3	450	2 I49481	alpha-2 adrenergic
21	86	2.3	2424	2 I46480	calcium channel BR
22	84.5	2.3	464	2 S23697	extensin - Volvox
23	84	2.2	424	2 A54964	spliceosome-associ
24	84	2.2	514	1 W2W15	B2 protein - human
25	84	2.2	539	2 T28770	hypothetical prote
26	84	2.2	603	2 T45093	hypothetical prote
27	83.5	2.2	282	2 F96687	hypothetical prote
28	83.5	2.2	437	2 S41766	heterogeneous nucl
29	83.5	2.2	684	2 T36771	probable integral

30	83.5	2.2	1215	2 T32734	myosin-IA - Acanth
31	83	2.2	309	2 T47685	probable RNA bindi
32	83	2.2	531	2 T08760	hypothetical prote
33	83	2.2	1106	2 U00405	hypothetical 119.5
34	83	2.2	1248	2 T05059	hypothetical prote
35	82.5	2.2	125	2 A71245	hypothetical prote
36	82.5	2.2	278	2 T46458	hypothetical prote
37	82.5	2.2	380	2 D70516	hypothetical prote
38	82.5	2.2	3345	2 T13423	hypothetical prote
39	82	2.2	312	2 A61183	hypothetical prote
40	82	2.2	344	2 T27410	hypothetical prote
41	82	2.2	704	2 A30411	synapsin Ia - rat
42	82	2.2	1414	1 S23809	collagen alpha 2(I
43	81.5	2.2	414	2 C86301	arginine/serine-ri
44	81.5	2.2	448	2 S05355	hypothetical prote
45	81.5	2.2	839	2 F75518	hypothetical prote

ALIGNMENTS

RESULT 1
C40201
artifact-warning sequence (translated ALU class C) - human
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 13-May-2000
C/Accession: C40201
R/Clavette, J.M.
Personal communication, 1992
A/Reference number: A40201
A/Accession: C40201
A/Molecule type: DNA
A/Residues: 1-613 <CIA>
R/Clavette, J.M.
Genomics 12, 838-841, 1992
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential
A/Reference number: A40200; PMID:92241891; PMID:1572661
A/Contents: annotation
C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C/Comment: Any significant similarity of a predicted protein sequence to a portion of t

Query Match
Best Local Similarity 43.9%; Pred. No. 0.00021;
Matches 43; Conservative 7; Mismatches 46; Indels 2; Gaps 2;

CY 687 GEXXLPVIALXXAGRSLSRRVDQXQHGRTSLQKTKIXVLAHCSLSYRSG 746
DB 2 GVCWLTTPVLPALMEAVTGSXQGEFTSLANNAKPHFYKRYKLAGHG-AGACSPYLAG 60

CY 747 XGXXIDACEVEAAVRDXAIALQPGXRETLISQKXK 784
DB 61 XGRMAXTRVELAVSR-DLAIALQPGXRETLISQKXK 97

RESULT 2
F40201
artifact-warning sequence (translated ALU class F) - human
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C/Accession: F40201
R/Clavette, J.M.
Personal communication, 1992
A/Reference number: A40201
A/Accession: F40201
A/Molecule type: DNA
A/Residues: 1-673 <CIA>
R/Clavette, J.M.
Genomics 12, 838-841, 1992
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential
A/Reference number: A40200; PMID:92241891; PMID:1572661
A/Contents: annotation

C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'X'.

Query Match 3.0%; Score 111; DB 4; Length 673;
Best Local Similarity 37.2%; Pred. No. 0.011; Mismatches 51; Indels 2; Gaps 1;
Matches 35; Conservative 6;

691 LIPVILALXXAXGSLRSRYVDPXQHGELSLQCKTKIXPXYLAFTCSLSYSEKXGX 750
6 LIPVILALMETRRAHLSRGVXDQDQHGELSLIKIKHLARN--GGPCLSLQHLGRQR 63

751 IDXAQVEAAVXRKXALALPQCKREFTLSQCKX 784
64 ENRLNPGXGCEPFLATLQPCQSKGLSKQK 97

RESULT 3

Q2B3 BHFL1 protein - human herpesvirus 4 (strain B95-8)

C/Species: human herpesvirus 4, Epstein-Barr virus
C/Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C/Accession: A03742

R/Banker, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A/Reference number: A93065; PMID:85035713; PMID:8092825

A/Accession: A03742

A/Molecule type: DNA

A/Residues: 1-660 <BAN>

R/Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984

A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A/Reference number: A03794; PMID:84270667; PMID:6087149

A/Contents: annotation; protein coding region

C/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52

C/Superfamily: human herpesvirus 4 BHFL1 protein

Query Match 2.8%; Score 104; DB 1; Length 660;
Best Local Similarity 23.2%; Pred. No. 0.053;
Matches 87; Conservative 13; Mismatches 103; Indels 172; Gaps 21;

18 AFGAP--PVSPRGARCGKRRPCG-----PVYKXYSDBXXTDPVRGGEPR- 61
188 AFGAPGTPAAR--GPGGAAVPSGATPBRGSGPADPPA-----AARLPPERQ--EPRL 238

62 -GALAYGAKRPAARRGATPSGXSARMGVYLGRTCTQXXOVSXGELREDN---LPVX 116
239 PDLAAARCPAC--PPPTSGMAA-----QTHRRPDPGPRRARRPGCPRTR 285

117 RRAKARLILFTNTDXESGASRFXPGFXAXXVKTGTTGLMRBSHSDVAFXSD 176
286 RR-----SGAQRGHPPPG-----ACQRPSPGTG--GRPAAPG----- 315

177 VGSYHKEAFTRKWTIVHPIIGXSWXIVVQVSTLLXMCCHNPQYERNRXRHL 236
316 -----APGTPA----- 321

237 VVVLGKANGAKXXSVGLXNLNASKSESRPXGTRRGASVGLGXFXPLSP----- 288
322 ----APGPGGAANVPSG-----ATPHERSGPADPPAARLPEROEPRLPDLAAARQC 373

289 PAGRPPSTKXXAAGGVPRAPG-----PGSXAECPSWE-----TGGRK 330
374 PAG--PPPTSGAAGQTHRRPPGCPSSAKNPG-----CPRTWRBSGAQRGHPPPAQR 427

331 -----GGXPLARNAP 340
428 PSGETGRPAAPGAP 442

RESULT 4

C75580 adenine deaminase-related protein - Deinetococcus radiodurans (strain R1)

C/Species: Deinetococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C/Accession: C75580

R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaleski, C.

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium Deinetococcus radiodurans R1.

A/Reference number: A75250; PMID:120036896; PMID:10567266

A/Accession: C75580

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-376 <WHI>

A/Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12376.1; PID:96460

A/Experimental source: strain R1

A/Genetic: DR0268

A/Map position: 2

Query Match 2.7%; Score 103; DB 2; Length 376;
Best Local Similarity 23.0%; Pred. No. 0.037;
Matches 82; Conservative 15; Mismatches 123; Indels 136; Gaps 18;

20 GADPVSPRGARCGKRRP--CGPVYKHYSDRXXTDPVRGGEPRGAL---AXGAKRPAAR 74
64 GRPAPAGAGGQSGAGBFRVVAHPALY--SGRADPAGAGAGAAARREILGPAPRQD 120

75 RPGATRS--GKSARWGVYLGRTCTQXXOVSXGELR-----EDRNLPR--WYRRAKRL 123
121 RPNVSRRLRRAARLATLRG-----AGVLRRGSGAARRDAPAPARRRRPRRL 169

124 ILFTNTDXESGASRFXPGFXAXXVKTGTTGLMRBSHSDVAFXSPVSSYX 183
170 -----GRHLS-----AALAHADVSPDRH----- 192

184 EAFTKXWIVHPIIGXSWXIVVQVSTLLXMCCHG---NPAQYERNRXXRLVTVL 240
193 -----RAGAGGRRRPAARRRRLRARRVVELL 219

241 GXGANGAKXXSVGLXNLNASKSESRPXGTRRGAS-----VGIXFXPLR 286
220 DVGRPARAPR--GHOSARRASGAPAR--RQRGPARAGSGARTARRRRRPGRRGPR 276

287 SPAPGRPPSTRKXARAGVPRAPGPGSXAECPSMETGRGKXGFLA--RHAP 340
277 VAPLR--GSDDRPAAGR--GRGCPGPD-----CGGALAGLHLSHHP 318

RESULT 5

B40201 artifact-warning sequence (translated Alt class E) - human

C/Species: Homo sapiens (man)
C/Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C/Accession: E40201

R/Claverie, J.M.

personal communication, 1992

A/Reference number: A40201

A/Accession: E40201

A/Molecule type: DNA

A/Residues: 1-597 <CLAS>

R/Claverie, J.M.

Genomics 12, 838-841, 1992

A/Title: Identifying coding exons by similarity search: Alu-derived and other potential

A/Contents: annotation

C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of

in-frame stop codons are shown as 'X'.

C/Comment: Any significant similarity of a predicted protein sequence to a portion of

Query Match 2.7%; Score 103; DB 4; Length 597;

R:Cheng, A.K.
Nucleic Acids Res 17, 4637-4646, 1989
A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies virus
A:Reference number: 504713; MUID:89315207; PMID:2346124
A:Accession: S04713
A:Molecule type: DNA
A:Residues: 1-1460 <CHE>
C:Superfamily: herpesvirus immediate-early protein 13175
C:Keywords: DNA binding; early protein; transcription regulation

Query Match 2.5%; Score 94; DB 1; Length 1460;
Best Local Similarity 20.8%; Pred. No. 1.2;
Matches 73; Conservative 18; Mismatches 122; Indels 138; Gaps 15;

```

QY 18 APGAPVSPRGARG-----GXRPPCPVKYHSDRXXTDPVRGQ-----58
DB 184 APPAPP-APRRPRDGGPRGSTRSVSPGRRRGLGPRRHQSQRW---PGRHGGGPLPQ 239
QY 59 --EPRGALAGAKRPAARPGATRSGXSAWGVXIGRYTCQTXQVSGELREDRLPMX 116
DB 240 PPPPPGR-----SRPPAAAPPPAE-----GTAVVTITSTAS-----PWL 274
QY 117 RRAARLLIFSTMTDSSGASRSPFXPRGXVXVRAVTTGITGLMPRSHSDVAFKSF 176
DB 275 -----DBPAAARLDP-----AAAPRPER-----234
QY 177 VGSYHXAEEFTKMTVHPLIGXXSWYTVVRQVSTLLMCCCHGNPAQYERNRXXRL 236
DB 295 -----LQDP-----QLQLD-----HHRRRRRARRP 315
QY 237 VYVLGXGAKGKXSVGLXMASSESPXGTIRORRG-ASVGLGXFPXPLSPAPGPP 295
DB 316 RPRRRGRTPRRGSGADLQROPFRRAEGEALTRGRFSSSGSDSLSP--ASPS 373
QY 296 STRXXRAGRVRRAPGPGSXAECPS-----WETGRKRGKXPLARHAP 340
DB 374 APRAPAAAAAARSASSSSSSSSSSSSSEGEDEGVAPGAPLARAGP 424

```

RESULT 10

A38316
alpha-2-adrenergic receptor - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 13-Aug-1999
C:Accession: A38316
R:Guyer, C.A.; Hozetman, D.A.; Wilson, A.L.; Clark, J.D.; Cragoe Jr., E.J.; Limbird, L.E.
J. Biol. Chem. 265, 17307-17317, 1990
A:Title: Cloning, sequencing, and expression of the gene encoding the porcine alpha-2-adrenergic receptor
A:Reference number: A38316; MUID:91009167; PMID:2170371
A:Accession: A38316
A:Molecule type: DNA
A:Residues: 1-450 <GV>
A:Cross-references: GB:J05652; NID:g164303; PIDN:AAA30984.1; PID:g164304
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein

Query Match 2.5%; Score 93; DB 2; Length 450;
Best Local Similarity 24.3%; Pred. No. 0.43;
Matches 43; Conservative 14; Mismatches 74; Indels 46; Gaps 8;

```

QY 241 GKGANGAKKXSVGLXMASSESPXGTIRORRGASVGLKXPLSPAPRPPSRXX 300
DB 258 GVGRRGAEAPLPVQLNGAPGEPAPAGP---RQAD-GLDUESSSSSHAEPRGEPHSE 312
QY 301 RA-----GGRVRR---APGSGXACPSWETGRKRGKXPLARHAPVR 343
DB 313 RGPRAKSKARASQVVRGDSLPFRGPGAGPGAPA-----TGAGEERGQ-----VA 357
QY 344 AADEFKXSTINRHTSACIFKXXILFLW-----VDIQKWDK---XTMKFYFMF 391
DB 358 KASRNRRGQRREKRFVLAIVIGVAVCMFPFFFTYTLTAVGCSVPPTLRFKFFWF 414

```

RESULT 11

G84348
hypothetical protein Vng1983h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84348
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.; Leibauser, B.; Keller, K.; Cruz, R.; Danon, M.U.; Hough, D.W.; Maddocks, D.G.; Jdb, Jung, K.H.; Alam, M.; Freilass, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; I.
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: G84348
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <STO>
A:Cross-references: GB:A804437; NID:g10581414; PIDN:AA620155.1; GSPDB:GN00138
C:Genetic: VNG1983h

Query Match 2.5%; Score 93; DB 2; Length 494;
Best Local Similarity 24.8%; Pred. No. 0.47;
Matches 102; Conservative 26; Mismatches 183; Indels 100; Gaps 24;

```

QY 1 ITHHEGETXEVXNKKXEAAGAPPVSPRGARGXR-PCGPPVKYHSDRXXTDPVRGGE 59
DB 62 LTPARGTGVDASADDAAPVPTVAVDGGGQTRCPSPGVSQVT---AEPVRR--R 116
QY 60 PRGALAXGAKRPAAR--RPGATR-----SGSARNG-----YXIGHYTCTQTXQ 101
DB 117 PVGRLLRVLTRGRARVDRAGVGRARADTASGPRRGETRPHADVCGCGPVD--R 173
QY 102 VEXGELREDRN-LPMXRPAKARLLIFST---NTDXSGASRSPFXPGXAXVRYKTYT 156
DB 174 VGANRPAEPRRLHCHCRLLRRLLVALQDALGRAGDRGHA-----VLPRRLAVP 225
QY 157 GITGL-----WPSXKSDVAFKSPDVG-SSYHXAEEFTKMTVHPLIG-----XX 200
DB 226 RVGRPVAVDLRPRGRDTR-PALDPGRTAADREPRPR--RPRVGGPRRPLVRPLR 281
QY 201 SWXITVVRQVSTLLMCCCHG-----NPAQYERNRXXHLYVLGXGANGAKKXSV 252
DB 282 HLYDRIPRCGRSLRVAARHVAVGGRDSAPRGAARRPRRGR-----GLAGARQGV 334
QY 253 -GLXMASSE--SREXGTIRORRGASVGLGXFPX-RLSPAPGPPS---TRXXRAG 304
DB 335 PGRRRRAGRLRLAARAAVAGRGHGFGLGRRRPARLDDPGRRPAPVADVSRRRRTRG 394
QY 305 ---RVRRAPGSGSXAECPSWETGRKRGKXPLA-----RHAPHYRAR 345
DB 395 RRRRPPORARRPGARQC-----RPSKALPLAVTGRPRGRSLAGTRAR 438

```

RESULT 12

D40201
artifect-warming sequence (translated ALU class D) - human
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: D40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: D40201
A:Molecule type: DNA
A:Residues: 1-579 <CIA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential.
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

```

Query Match      2.4%; Score 90.5; DB 4; Length 579;
Best Local Similarity 35.2%; Pred. No. 0.98;
Matches 31; Conservative 9; Mismatches 33; Indels 15; Gaps 4;

QY 701 AKAGRSLESPRDPDPOXHGFTLSQ-KNTKIX-----PYVLAHTCSLSYSEKGGXIXD 753
    15 AMAGVSSSGVQEPGQVHETPLPKDKTISQWKMKVYVPA-----TKAGAGSLDI 69
    754 AQEVEAAVFXDXALALQPKXRETTLSQ 781
Db 70 GSQ---GSVSHDTSVLPQGHKTLP 94

RESULT 13
T45134
hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
C:Species: Microbacterium ammoniaphilum
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45134
R:Striebel, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.
Gene 172, 41-46, 1996
A:Title: Cloning and characterization of the ManI restriction-modification system from M
A:Reference number: 222923; MUID:96257250; PMID:8654988
A:Accession: T45134
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-529 <STR>
A:Cross-references: EMBL:X79027; NID:984667; PIDN:CA55649.1; PID:91679831
A:Experimental source: ATCC 15354

Query Match      2.4%; Score 90; DB 2; Length 529;
Best Local Similarity 21.9%; Pred. No. 1;
Matches 87; Conservative 23; Mismatches 140; Indels 148; Gaps 18;

QY 2 RHEHGEETXEV---XNKKEAPGAP---PVSFRGARGXRRCPG---PVKHYSDRXXTD 52
    70 RHRRPAQRQVHPLOPRHPEGARGELFVRDRAERGERREPAPARDARHL--RLGAM 127
    53 FVRSG-----GE-----PRGALXGAXRPPARRPPGATRSXGSRMG 88
    128 PARQVLRHRRHARGBRGPRQVPRQHPRG-----RRDAAGRPGLHRRRRAR-- 179
    89 VXLGRYTCQTXQVSGXG--ELREDPNLPMXRAARLILISTITDSESGASRFXFGF 146
    180 -----RQGRFPQVHGDQHRADPRRPRDPRAGHNP-----REGGQGEARPL-- 223
    147 XAXXVRKVTGTLGLMRPSXHSDDVAFXSFDVGSSTYHAEFTKEMIVHPLIGXSWXTV 206
    224 -----PARGGQGGGSGR--HPLGD-----HPLGD----- 244
    207 VRQVFTLLXMCCHGNPAQYERNRKHLYVTLCXGANGKXXSVGLXLANASSES--- 263
    245 -----EDRRARQ-----GAFAHRBALHILQRRGRADQC 276
    264 REXGTIRORRGASVGLGXPRLSPAPRPP-----STRXXRG-----RVP 307
    277 RAQGAARRARRRGGG-GLPREDRVRAHRRPRRGGGAAALDRAGVGARPARRRRLRH 335
    308 RRAFGSGXAECPSSMETGRGRKGGPLARHAPVHRR 345
    336 RPADLPFGGPGGVARLDHPQLGGAAGSRHRRHRLRR 373
Db

RESULT 14
T51112
beta-glucosidase [EC 3.2.1.21] oler [validated] - Streptomyces antibioticus (ATCC 11891)
C:Species: Streptomyces antibioticus
A:Variety: ATCC 11891
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T51112

```

R:Quiroz, L.M.; Aguirrezabalaga, I.; Olano, C.; Mendez, C.; Salas, J.A.

submitted to the EMBL Data Library, September 1999

A:Reference number: 225300

A:Accession: T51112

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-769 <QUT>

A:Cross-references: EMBL:AF055579; PIDN:AA012650.1

A:Experimental source: ATCC 11891

C:Genetics:

A:Gene: oler

C:Function:

A:Description: extracellular reactivation of oleandomycin; converts glycosylated oleam

C:Keywords: glucosidase; hydrolase

```

Query Match      2.4%; Score 89.5; DB 2; Length 769;
Best Local Similarity 29.1%; Pred. No. 1.7;
Matches 34; Conservative 8; Mismatches 48; Indels 27; Gaps 4;

QY 237 VYTLGXGANGAKXSVGLXLANASKESEPRXGTRRGA-----SVG---LGXEXPLS 287
    334 IAVIGPTAKVPKVTGIGSSYIVPDGASAPLDTRRAGAGSTRVYSTGBETVGVVPO-S 392
    288 PRAGRPPESTRXXRAG-----RVPRAPPGSGXACPPSSMETGR 327
    393 ALPRPRPBGVFPAGCGVLYDTITVYVGSYRIARAGGNAYVELDQGEPRGR 449
    393 ALPRPRPBGVFPAGCGVLYDTITVYVGSYRIARAGGNAYVELDQGEPRGR 449
Db

RESULT 15
S54570
probable membrane protein YMR109w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YW9718.08
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 02-Feb-2001
C:Accession: S54570
R:Hunt, S.; Bowman, S.
Submitted to the EMBL Data Library, May 1995
A:Reference number: S54510
A:Accession: S54570
A:Molecule type: DNA
A:Residues: 1-1219 <HUN>
A:Cross-references: EMBL:249702; NID:9817859; PID:9817867; MIPS:YMR109w
A:Experimental source: strain AB972
C:Genetics:
A:Gene: YMR109
A:Cross-references: SGD:S0004715; MIPS:YMR109w
A:Map position: 13R
C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homo;
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:39-702/Domain: myosin motor domain homology <MMOT>
F:129-136/Region: nucleotide-binding motif A (P-loop)
F:293-309/Domain: transmembrane #status predicted <TMW>
F:1092-1142/Domain: SH3 homology <SH3>

Query Match      2.4%; Score 89.5; DB 2; Length 1219;
Best Local Similarity 26.8%; Pred. No. 2.7;
Matches 30; Conservative 11; Mismatches 44; Indels 27; Gaps 4;

QY 256 INASKESEPPXGTRR-----GASVGLGXPRFLSPAPRPPSTTRXXRAGRV 307
    1035 LTNSGNSARSPPTATPATPATPATPAAMGSGRQA-NIPPPPPPPSPSKPEPEFEAA 1093
    308 RRAFGSGXAE-----PSGMEGR-----GRKGGXPLARHAP 341
    1094 YDFPGSGSPBELPLKKGDIYITREPPSGMSKGLDGSKEGVNPTANKPH 1145
Db

Search completed: April 27, 2004, 10:56:03
Job time : 18.903 secs

```